

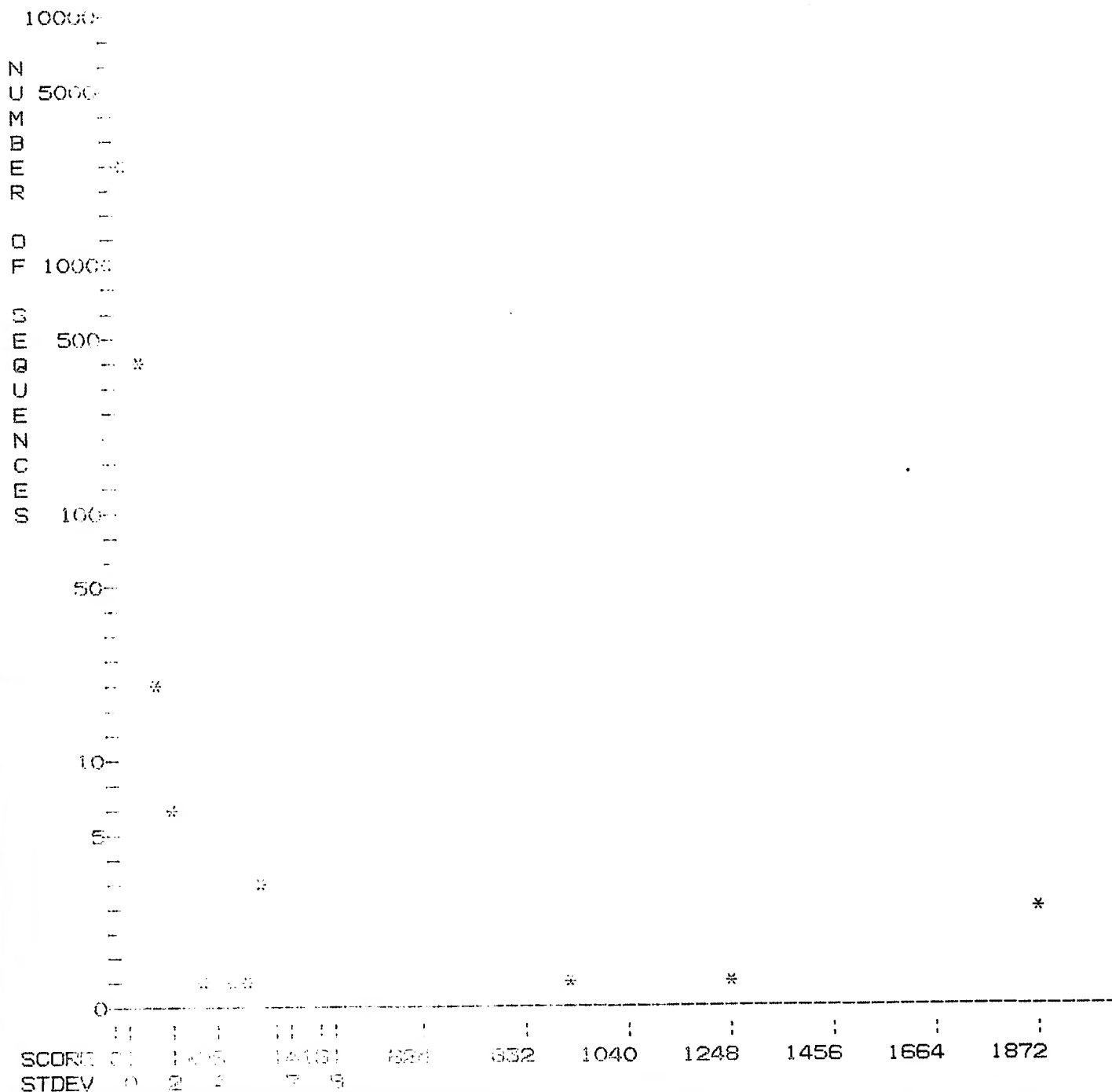
Kun3
07/158652
Claim 320

Results file: kunz-158-cl32umbl.res made by sheppard on Thu 8 Mar 90 9:51:34-PST.

EMBL

Query sequence being compared: KUNZ-158-CL32.SEQ
Number of sequences searched: 3460
Number of scores above cutoff: 10

Results of the initial comparison of KUNZ-158-CL32.SEQ with:
Data bank: UMBL 21, all entries



PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	147		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	40	38	46.71
Times:	CPU	Total Elapsed	
	00:07:21.04	00:21:31.00	

Number of residues:	5125998
Number of sequences searched:	3460
Number of scores above cutoff:	10

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 38 standard deviations above mean ****						
1. REHTLV3	Human T-cell leukaemia type II	9748	1872	2186	39.22	0
2. HIVH300	Human T-cell lymphotropic virus	9749	1872	2186	39.22	0
**** 25 standard deviations above mean ****						
3. HIVEL006	Human lymphadenopathy virus (E	9176	1246	1904	25.82	0
**** 18 standard deviations above mean ****						
4. HIVMAL06	Human lymphadenopathy virus (M	9229	916	2060	18.76	0
**** 5 standard deviations above mean ****						
5. HIV2R00X	Human immunodeficiency virus t	9671	306	1194	5.70	0
6. REHTLV40	STLV-3 (HTLV-4) partial provir	5391	299	1257	5.55	0
7. RESIVAXX	Simian immunodeficiency virus	9264	294	1250	5.44	0
8. RESIVM01	Simian immunodeficiency virus	9646	290	1247	5.35	0
**** 4 standard deviations above mean ****						
9. RESIV291	Simian immunodeficiency virus	1142	259	565	4.69	0
**** 3 standard deviations above mean ****						
10. H15127	Figure 1. Structure of the art	306	184	298	3.08	0

The scores below are sorted by optimized score.
Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. REHTLV3	Human T-cell leukaemia type II	9748	1872	2186	0.00	0
2. HIVH300	Human T-cell lymphotropic virus	9749	1872	2186	0.00	0

3.	HIVMAL03	Human lymphadenopathy virus (M	8229	916	2060	0.00	0
4.	HIVEL03	Human lymphadenopathy virus (E	9176	1246	1904	0.00	0
5.	REHTLV40	HTLV-3 (HTLV-4) partial provir	5391	299	1257	0.00	0
6.	REHTLV50	Human immunodeficiency virus	9264	294	1250	0.00	0
7.	REHTLV60	Human immunodeficiency virus	9646	290	1247	0.00	0
8.	HIVERS04	Human immunodeficiency virus t	9671	306	1194	0.00	0
9.	REHTLV70	Human immunodeficiency virus	1142	259	565	0.00	0
10.	MIS127	Figure 1. Structure of the art	306	184	298	0.00	0
SEARCH STATISTICS							

Scores:	Mean	Median	Standard Deviation
	2186	2187	0.00

Times:	CPU	Total Elapsed
	00:00:00.00	00:00:00.00

Number of residues:	73322
Number of sequences optimized:	10

1. KUNZ-158-CL3A.SEB

REHTLV3 Human T-cell leukaemia type III (HTLV-III) provira

ID REHTLV3 standard; RNA; 8748 BP.

XX

AC X017621

XX

DT 03-SEP-1987 (an correction)

DT 01-SEP-1987 (an correction)

DT 03-AUG-1987 (an correction)

DT 22-OCT-1986 (minor modification)

DT 05-NOV-1985 (RV added)

DT 26-MAR-1985 (first entry)

XX

DE Human T-cell leukaemia type III (HTLV-III) proviral genome

DE (AIDS virus for acquired immune deficiency syndrome)

XX

KW acquired immune deficiency syndrome; direct repeat; endonuclease;

KW glycoprotein; inverted repeat; protease; provirus;

KW reverse transcriptase; terminal repeat.

XX

OS Human T-cell leukemia virus type III

OC Viridae; ss-RNA enveloped viruses; Retroviridae.

XX

RN [1] (bases 1-8748)

RA Ratner L., Haseltine W., Patarca R., Livak K. J., Starcich B. R.,

RA Josephs S. F., Doran E. R., Rafalski J. A., Whitehorn E. A.,

RA Baumeister K., Ivanoff L., Pettaway S. R. Jr., Pearson M. L.,

RA Lautenberger J. A., Papas T. S., Ghayeb J., Chang N. T., Gallo R. C.,

RA Wang-staal F.

RT "Complete nucleotide sequence of the AIDS virus, HTLV-III";

RL Nature 313:277-284(1985).

XX

RN [2]

RA Muesing M. A., Smith D. H., Cabradilla C. D. JR., Benton C. V.,

RA Kasky L. A., Capon D. J.;

RT "Nucleic acid structure and expression of the human AIDS/

RT lymphadenopathy retrovirus";

RL Nature 313:450-458(1985).

XX

FH	Key	From	To	Description
----	-----	------	----	-------------

FH

FT	INVREP	1	2	inverted repeat
----	--------	---	---	-----------------

FT	SITE	1	824	long terminal repeat
----	------	---	-----	----------------------

FT	PRM	127	430	TATA-box
----	-----	-----	-----	----------

FT	SITE	453	453	U3 region
FT	CAP	454	454	cap site
FT	SITE	454	551	R region
FT	SITE	552	634	U5 region
FT	INVREP	633	634	inverted repeat
FT	SITE	635	653	tRNA binding site (tRNA-Lys)
FT	CDS	787	1182	gag p17
FT	CDS	787	2321	gag precursor polypeptide
FT	CDS	1183	2321	gag p24 and gag p15 for
FT				major capsid protein and for
FT				put. retroviral nucleic acid
FT				binding protein (NBP)(ref.2)
FT				(boundaries not defined)
FT	RPT	1968	2002	direct repeat
FT	RPT	2031	2065	direct repeat
FT	CDS	2081	5125	pol precursor polypeptides
FT				put. protease at 5' terminus
FT				reverse transcriptase
FT				put. endonuclease at 3' terminus
FT	RPT	2128	2163	direct repeat
FT	RPT	2164	2176	direct repeat
FT	CDS	5040	5648	SDR short open reading frame
FT				pot. vestigial env gene
FT	CDS	6323	8821	env-lor precursor polypeptide
FT	CDS	6323	8821	envelope glycoprotein
FT	SITE	7786	7787	put. peptide cleavage site
FT	CDS	7787	8821	put. lor transmembrane
FT				protein
FT	SITE	9098	9103	poly purine stretch
FT	SITE	9115	9567	U3 region
FT	RPT	9116	9748	long terminal repeat
FT	SITE	9580	9665	R region
FT	SITE	9641	9646	polyadenylation signal
FT	SITE	9689	9748	U5 region
FT	INVREP	9747	9748	inverted repeat
XX				
SS	Sequence	9748 BP;	3431 A; 1781 C; 2368 G; 2168 T; 0 other;	

[illegible]

[illegible]

AC K02010; K02008; K02009;
 XX
 DT 08-SEP-1987 (an correction)
 DT 03-SEP-1987 (an correction)
 DT 02-SEP-1987 (an correction)
 DT 02-SEP-1987 (an correction)
 DT 01-SEP-1987 (an correction)
 DT 23-JUN-1987 (minor modifications)
 DT 29-OCT-1986 (incorporated)
 XX
 DE Human T-cell lymphotropic virus type III, complete reference genome
 DE (isolates HXB2, HXB3, BH10, BH5 and BH8 of HTLV-III DNA).
 XX
 KW acquired immune deficiency syndrome; complete genome; env gene;
 KW gag gene; long terminal repeat; pol gene; polyprotein; provirus;
 KW reverse transcriptase; TAR protein; trans-activator.
 XX
 OS Human T-cell lymphotropic virus type III
 OC Viridae; ss-RNA enveloped viruses; Retroviridae.
 XX
 RN [1] (bases 1-853, 9113-9749)
 RA Starcich B.R., Ratner L., Josephs S.F., Okamoto T., Gallo R.C.,
 RA Wong-staal F. ;
 RT "Characterization of long terminal repeat sequences of htlv-iii";
 RL Science 227:538-540(1985).
 XX
 RN [2] (bases 1-9749)
 RA Ratner L., Haseltine W., Patarca R., Livak K.J., Starcich B.R.,
 RA Josephs S.F., Doren E.R., Rafalski J.A., Whitehorn E.A.,
 RA Baumeister K., Ivanoff L., Petteway S.R. Jr., Pearson M.L.,
 RA Lautenberger J.A., Papas T.S., Ghrayeb J., Chang N.T., Gallo R.C.,
 RA Wong-staal F. ;
 RT "Complete nucleotide sequence of the AIDS virus, HTLV-III";
 RL Nature 313:277-284(1985).
 XX
 RN [3] (bases 508-8668) exons only, tat mrna
 RA Arya S.K., Guo C., Josephs S.F., Wong-staal F. ;
 RT "Trans-activator gene of human T-lymphotropic virus type III
 RT (HTLV-III)";
 RL Science 220:68-73(1985).
 XX
 RN [4] (bases 5775-6082, 8397-8493)
 RA Sodroski J.G., Patarca R., Rosen C.A., Wong-staal F., Haseltine W. ;
 RT "Location of the trans-activating region on the genome of human
 RT T-cell lymphotropic virus type III";
 RL Science 229:74-77(1985).
 XX
 RN [5] mrna splice sites
 RA Rabson A.B., Laugherty D.F., Venkatesan S., Boulukos K.e.,
 RA Benn S.L., Folks T.M., Feurino P., Martin M. ;
 RT "Transcription of novel open reading frames of AIDS retrovirus
 RT during infection of lymphocytes";
 RL Science 223:1388-1390(1985).
 XX
 RN [6] 27k antigen cds
 RA Allan J.S., Coligan J.E., Lee T.H., McLane M.F., Kanki P.J.,
 RA Gimpson J.E., Essex M. ;
 RT "A new HTLV-III/LAV encoded antigen detected by antibodies from
 RT AIDS patients";
 RL Science 230:310-313(1985).
 XX
 RN [7] (bases 5772-8933) in hxb-3
 RA Crowl R., Ganguly K., Gordon M., Conroy R., Schaber M., Kramer R.,
 RA Shaw S., Wong-staal F., Reddy R.P. ;
 RT "HTLV-III env gene products synthesized in E.coli are recognized
 RT by antibodies present in the sera of AIDS patients";

RL Cell 41:979-983(1985).
 XX
 RN 141 gp160 and gp120 coding sequences
 RA Allen J.S., Coligan J.E., Barin F., McLane M.F., Sodroski J.G.,
 RA Rosen C.A., Haseltine W.A., Lee T.H., Essex M.;
 RT "Major glycoprotein antigens that induce antibodies in AIDS
 RT patients are encoded by HTLV-III";
 RL Science 228:1031-1034(1985).
 XX
 RN 191 regulatory sequences in the ltr
 RA Rosen C.A., Sodroski J.G., Haseltine W.A.;
 RT "The location of cis-acting regulatory sequences in the human T
 RT cell lymphotropic virus type III (HTLV-III/LAV) long terminal
 RT repeat";
 RL Cell 41:813-823(1985).
 XX
 RN [101] (bases 1-9749)
 RA Van Beveren C., Van Beveren C., Van Beveren C.;
 RT "Appendix B: HTLV-3/LAV genome";
 RL (in) Weiss R., Teich N., Varmus and Coffin J.M. (eds.);
 RL RNA Tumor Viruses Second Edition:1102-1148
 RL Cold Spring Harbor Laboratory, New York (1985)
 XX
 RN [111] trans-activator function and tar sequence
 RA Rosen C.A., Sodroski J.G., Goh W.C., Dayton A.I., Lippke J.,
 RA Haseltine W.A.;
 RT "Post-transcriptional regulation accounts for the trans-activation
 RT of the human T-lymphotropic virus type III";
 RL Nature 318:555-559(1986).
 XX
 RN [113] pol coding sequence
 RA Marco Veronesi F., Copeland T.D., DeVico A.L., Rahman R.,
 RA Omazian S., Gallo R.C., Sarngadheran M.G.;
 RT "Characterization of highly immunogenic p66/p51 as the reverse
 RT transcriptase of HTLV-III/LAV";
 RL Science 231:1209-1251(1986).
 XX
 RN [131] the 23k sor gene product
 RA Ken N.C., Franchini G., Wong-staal F., DuBois G.C., Robey W.G.,
 RA Lautenberger J.A., Papas T.S.;
 RT "Identification of HTLV-III/LAV sor gene product and detection of
 RT antibodies in human sera";
 RL Science 231:1553-1555(1986).
 XX
 RN [141] pol nh2-terminal region
 RA Kremer R.A., Schaber M.D., Skalka A.M., Ganguly K., Wong-staal F.,
 RA Reddy R.P.;
 RT "HTLV-III gag protein is processed in yeast cells by the virus
 RT pol-protease";
 RL Science 231:1580-1584(1986).
 XX
 RN [151] sor 23k protein
 RA Lee T.H., Coligan J.E., Allen J.S., McLane M.F., Groopman J.E.,
 RA Essex M.;
 RT "A new HTLV-III/LAV protein encoded by a gene found in cytopathic
 RT retroviruses";
 RL Science 231:1546-1549(1986).
 XX
 RN [161] sor 23k protein
 RA Sodroski J.G., Goh W.C., Rosen C.A., Tartar A., Portetelle D.,
 RA Buny A., Haseltine W.;
 RT "Replicative and cytopathic potential of HTLV-III/LAV with sor
 RT gene deletions";
 RL Science 231:1548-1553(1986).
 XX
 RN [171] 5' flanking sites in the promoter region

RA James K.A., Kankana J.T., Luciw P.A., Tjian R.;
 RT "Activation of the AIDS retrovirus promoter by the cellular
 RT transcription factor, Sp1";
 RL Science 232:755-759(1986).
 XX
 RN [10] acceptor and donor splice sites for tat and 27k
 RA Anya S.K., Gallo R.C.;
 RT "Three novel genes of human T-lymphotropic virus type III: Immune
 RT reactivity of their products with sera from acquired immune
 RT deficiency syndrome patients";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:2209-2213(1986).
 XX
 RN [11] deletion mutants in the tat gene
 RA Dayton A.L., Sodroski J.B., Rosen C.A., Goh W.C., Haseltine W.A.;
 RT "The trans-activator gene of the human T cell lymphotropic virus
 RT type III is required for replication";
 RL Cell 44:841-847(1986).
 XX
 RN [12] hypervariable and conserved regions in the env gene
 RA Willey R.W., Ruthledge R.A., Dias S., Folks T., Theodore T.S.,
 RA Buckler C.E., Martin M.A.;
 RT "Identification of conserved and divergent domains within the
 RT envelope gene of the acquired immunodeficiency syndrome
 RT retrovirus";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5038-5042(1986).
 XX
 RN [21] env cds boundaries
 RA Sodroski J.B., Goh W.C., Rosen C.A., Dayton A., Terwilliger E.,
 RA Haseltine W.;
 RT "A second post-transcriptional trans-activator gene required for
 RT HTLV-III replication";
 RL Nature 321:412-417(1986).
 XX
 CC Sequence for [7] was kindly supplied in computer readable form by
 CC R. Crowl, 09/17/85. R. Patarca provided sites information and a
 CC clean copy for [4], 09/16/85. Acquired immune deficiency syndrome
 CC (AIDS) is caused by a retrovirus known by several names, perhaps
 CC representing two separate strains: human T-cell lymphotropic
 CC virus-III (HTLV-III), whose sequence is given below, and
 CC lymphadenopathy-associated virus (LAV) are thought to be one strain
 CC differing from AIDS-associated retrovirus type 2 (ARV-2) when
 CC overall homology is the criterion. Some reading frame similarities
 CC suggest that ARV-2 and LAV are more closely related. All three
 CC viruses, whose sequences do not differ by more than 6%, are
 CC believed to belong to the C type subfamily Lentiviridae, the "slow"
 CC retroviruses. The BH10 sequence differs from BH8 and BH5 by 0.9% in
 CC the coding regions and 1.3% in the noncoding regions, and the
 CC authors of [2] believe that these are stable variants. The 5' and
 CC 3' LTRs of BH10 and BH8 were not fully sequenced; the missing bases
 CC (493-675 and 8608-8749) were filled in by [2] from the proviral
 CC clone HXB2 [1]. The sequence below is that of BH10 with exception
 CC of the variation at position 5197 which allows annotation of the
 CC 27K coding sequence. The BH8 sequence spans bases 6033 to 9607, the
 CC BH5 sequence spans bases 675 to 6038, and the HXB3 sequence [7]
 CC spans bases 5778 to 9883. While this entry is offered as the
 CC reference locus for the AIDS retroviral sequence loci, no claim is
 CC being made that this sequence is more prevalent or typical than
 CC others, all of which have been entered in this library with
 CC annotation. The HTLV-III genome encodes at least six proteins or
 CC polyproteins: gag, pol, env, TAT, 27K antigen and the sor 23K
 CC product. The 3' ORF (positions 8797-9447) is truncated in BH10
 CC (stop codon at positions 8196-8198), but reads through in BH8 and
 CC other sequences to yield what is now called the 27K antigen. The
 CC sequence below is from BH10 with exception of the variation at
 CC position 5197 which allows annotation of the 27K coding sequence.
 CC Additionally there are four short open reading frames, bases

CC 1943-1948, 4442-4642, 5532-5926 and 6035-6340, which are conserved
 CC to a large degree. A seventh gene has been proposed based upon a
 CC combination of mutational and regulatory evidence: called "ART" (
 CC for anti-repression transactivator), its product appears to act
 CC post-transcriptionally to relieve negative repression of gag and
 CC env production [21]. The exon assignments for ART are putative, but
 CC if they are corroborated, the ART protein would be 116 amino acids
 CC in length. The mechanism for pol gene translation has not been
 CC elucidated: a gag-pol fusion protein is possible; splicing or
 CC frameshift have not been ruled out. The viral protease would be
 CC determined by the region in question. Approximately two-thirds of
 CC the variant sites in the gag and pol genes are "silent mutations",
 CC while over half of those in the env gene are not. Reference [20]
 CC defines divergent and conserved regions for the env gene. Because
 CC of the excessive variability of the env gene, differences between
 CC the sequences summarized herein and other env gene entries have not
 CC been associated; only HTLV-III sequence variations have been
 CC included in the sites of this entry. Other entries will include
 CC information for alignment with this entry, including the Zaire and
 CC New York isolate sequences reported by [20]. The TAT protein
 CC (trans-activator protein, approximately 14 kd) is an effector of an
 CC autostimulatory pathway through interaction with a positive control
 CC element, the trans-activating responsive sequence, TAR. TAT seems
 CC to be a transcriptional control molecule in HTLV-I, but [11]
 CC demonstrates that it is a post-transcriptional regulatory molecule
 CC in HTLV-III. Deletion mutants in the TAT gene are incapable of
 CC prolific replication and exhibit no cytopathic effects in T4+ cell
 CC lines [19]. The TAR sequence(s) are found to be between -17 and +80
 CC relative to the cap site +1 (base 455) and is highly conserved.
 CC Enhancer sequences which need not be viral-specific are found
 CC upstream from TAR [8] [11]. Three tandem decanucleotide Sp1 binding
 CC sites are located between bases 377 and 409, of which site III
 CC shows the strongest affinity for the cellular factor; intact, the
 CC three sites cause up to a tenfold effect on transcriptional
 CC efficiency in vitro ([17] (The authors demonstrate the existence of
 CC SFI in a human T-cell line). In addition to the ~9.4 kb genomic
 CC mRNA, subgenomic mRNAs of 7.4, 5.5, 5.0, 4.3, 2.0 and 1.8 have been
 CC detected. All are probably polyadenylated at the same site,
 CC position 5866 below, with a potential polyadenylation signal at
 CC 5042-5545 and capped at the same site, position 455, with a
 CC potential TATA box at 427-431. The doubly-spliced transcript of
 CC about 2.0 kb is responsible for the TAT message at least, and
 CC depending upon the acceptor site, also for the sor and 27K
 CC messages, given that a single, albeit partial, mRNA exists for all
 CC three [19]. The acceptor splice for TAT is at position 5811 and the
 CC putative acceptor splice for 27K is at position 6010; the donor
 CC splice site in all three cases would be at position 6079 [18]. The
 CC doubly spliced message would also encode the newly proposed ART
 CC protein.

XX

FH	Key	From	To	Description
FH				
FT	RPT	1	634	5' LTR
FT	RPT	1	634	5' LTR
FT	VARIANT	82	82	a in BH10; g in H9
FT	VARIANT	101	101	g in BH10; a in H9
FT	VARIANT	108	108	a in [2], H9; g in HXB2 [1]
FT	VARIANT	164	164	g in [2]; t in HXB2 [1], H9
FT	VARIANT	166	166	t in [2]; g in HXB2 [1], H9
FT	VARIANT	176	176	a in [2]; g in HXB2 [1], H9
FT	VARIANT	183	183	c in [2], H9; t in HXB2 [1]
FT	VARIANT	227	227	a in [2], H9; g in HXB2 [1]
FT	VARIANT	281	281	a in [2]; g in HXB2 [1], H9
FT	VARIANT	333	333	c in [2]; t in HXB2 [1], H9
FT	SITE	377	396	Sp1 binding site III [17]
FT	SITE	399	399	Sp1 binding site II [17]

FT	SITE	389	403	Sp1 binding site I [17]
FT	VARIANT	421	421	c in BH10, BH5; t in H9
FT	RPT	454	551	R repeat 5' copy
FT	RPT	454	551	R repeat 5' copy
FT	PROVRL	454	5656	HIVLV3 virion RNA
FT	CAP	455	455	genomic mRNA start (cap site) [10]
FT	CAP	455	455	TAT, ART mRNA exon 1 start (cap site)
FT				[10], [18], [21]
FT	VARIANT	501	501	a in BH10, BH5, H9; g in HXB2 [1]
FT	SITE	636	653	primer (Lys-tRNA) binding site
FT	VARIANT	654	654	c in BH10, BH5; t in H9
FT	VARIANT	677	677	g in BH10, BH5; ggag in H9
FT	VARIANT	704	704	tga in BH10, H9; g in BH5 [2]
FT	CDS	787	2325	gag polyprotein precursor
FT	VARIANT	1290	1290	a in BH10; g in BH5 [2], H9
FT	VARIANT	1431	1431	a in BH10; g in BH5 [2], H9
FT	VARIANT	1455	1455	t in BH10, H9; c in BH5 [2]
FT	VARIANT	1611	1611	a in BH10, H9; g in BH5 [2]
FT	VARIANT	1620	1620	c in BH10, H9; t in BH5 [2]
FT	VARIANT	1650	1650	a in BH10, H9; g in BH5 [2]
FT	VARIANT	1662	1662	t in BH10; c in BH5 [2], H9
FT	VARIANT	1675	1675	g in BH10, BH5; c in H9
FT	VARIANT	1722	1722	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1805	1805	g in BH10, BH5; a in H9
FT	VARIANT	1845	1845	a in BH10, BH5; g in H9
FT	VARIANT	1903	1903	a in BH10, H9; t in BH5 [2]
FT	VARIANT	1905	1905	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1923	1923	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1950	1950	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1953	1953	g in BH10, H9; t in BH5 [2]
FT	VARIANT	1983	1988	c in BH10, H9; t in BH5 [2]
FT	VARIANT	1992	1992	c in BH10, H9; a in BH5 [2]
FT	VARIANT	2003	2003	g in BH10, H9; a in BH5 [2]
FT	VARIANT	2013	2013	g in BH10, H9; a in BH5 [2]
FT	CDS	2391	5129	pol polyprotein (NH2-terminus
FT				uncertain; AA at 2391)
FT	VARIANT	2468	2468	g in BH10, BH5; a in H9
FT	VARIANT	2581	2591	c in BH10, H9; t in BH5 [2]
FT	VARIANT	2600	2600	g in BH10, H9; a in BH5 [2]
FT	VARIANT	2741	2741	g in BH10; a in BH5 [2], H9
FT	VARIANT	2827	2827	a in BH10, H9; g in BH5 [2]
FT	VARIANT	2856	2856	a in BH10, H9; g in BH5 [2]
FT	VARIANT	2860	2860	c in BH10, H9; t in BH5 [2]
FT	VARIANT	3007	3007	tta in BH10, H9; gtg in BH5 [2]
FT	VARIANT	3087	3087	a in BH10; g in BH5 [2], H9
FT	VARIANT	3124	3122	c in BH10, H9; t in BH5 [2]
FT	VARIANT	3232	3222	c in BH10, H9; t in BH5 [2]
FT	VARIANT	3302	3302	ag in BH10, H9; ga in BH5 [2]
FT	VARIANT	3358	3358	g in BH10, H9; a in BH5 [2]
FT	VARIANT	3383	3383	g in BH10, BH5; a in H9
FT	VARIANT	3385	3385	c in BH10, H9; t in BH5 [2]
FT	VARIANT	3755	3755	a in BH10, BH5; g in H9
FT	VARIANT	3767	3767	g in BH10, H9; a in BH5 [2]
FT	VARIANT	3833	3833	t in BH10, BH5; c in H9
FT	VARIANT	3855	3855	t in BH10, BH5; c in H9
FT	VARIANT	3858	3858	c in BH10, BH5; t in H9
FT	VARIANT	3822	3822	a in BH10, H9; g in BH5 [2]
FT	VARIANT	3934	3934	a in BH10, BH5; g in H9
FT	VARIANT	3954	3954	g in BH10, BH5; c in H
FT	VARIANT	3962	3962	caa in BH10, H9; tag in BH5 [2]
FT	VARIANT	3977	3977	g in BH10, H9; a in BH5 [2]
FT	VARIANT	3984	3984	c in BH10, H9; a in BH5 [2]
FT	VARIANT	3993	3993	a in BH10, H9; c in BH5 [2]
FT	VARIANT	4010	4010	a in BH10; g in BH5 [2], H9
FT	VARIANT	4016	4016	g in BH10, H9; a in BH5 [2]
FT	VARIANT	4029	4029	t in BH10, H9; c in BH5 [2]

FT	VARIANT	4043	4043	a in BH10; g in BH5 [2], H9
FT	VARIANT	4054	4054	c in BH10, H9; t in BH5 [2]
FT	VARIANT	4116	4116	a in BH10, BH5; c in H9
FT	VARIANT	4167	4167	g in BH10, BH5; c in H9
FT	VARIANT	4232	4232	t in BH10, H9; a in BH5 [2]
FT	CDS	5074	5052	sur 23K protein
FT	VARIANT	5156	5152	a in BH10, H9; g in BH5 [2]
FT	VARIANT	5314	5314	t in BH10, BH5; c in H9
FT	VARIANT	5343	5348	a in BH10, H9; g in BH5 [2]
FT	VARIANT	5401	5401	t in BH10, H9; c in BH5 [2]
FT	VARIANT	5412	5412	c in BH10, H9; t in BH5 [2]
FT	VARIANT	5543	5548	a in BH10, H9; g in BH5 [2]
FT	VARIANT	5628	5628	g in BH10, H9; a in BH5 [2]
FT	VARIANT	5846	5848	g in BH10, H9, HXB3; a in BH5 [2]
FT	CDS	5854	5078	TAT protein, exon 2 (first expressed exon)
FT	VARIANT	5934	5934	a in BH10, H9, HXB3; c in BH5 [2]
FT	CDS	6003	6078	ART protein, exon 2 (first expressed exon; putative)
FT	VARIANT	6035	6045	cctcctcaagg in BH10, HXB3 [7];
FT				gctcatcgaag in BH8 [2];
FT				g in BH5 [2], clone 12 cDNA [21]
FT	VARIANT	6086	6086	g in BH10, BH8, H9; a in HXB3 [7]
FT	VARIANT	6096	6096	t in BH10, HXB3 [7], H9; c in BH8 [2]
FT	VARIANT	6108	6108	a in BH10, HXB3 [7], H9; c in BH8 [2]
FT	VARIANT	6113	6114	gc in BH10, HXB3 [7], H9;
FT				gtaac in BH8 [2]
FT	VARIANT	6124	6124	a in BH10, HXB3 [7], H9; c in BH8 [2]
FT	VARIANT	6152	6152	g in BH10, HXB3 [7], BH8; c in H9
FT	CDS	6255	6225	envelope protein precursor (env)
FT	VARIANT	6373	6373	a in BH10, HXB3 [7], H9; t in BH8 [2]
FT	VARIANT	6474	6474	t in BH10, BH8 [2], H9; g in HXB3 [7]
FT	VARIANT	6748	6748	t in BH10, HXB3 [7], H9; a in BH8 [2]
FT	VARIANT	6929	6929	t in BH10, HXB3 [7], H9; c in BH8 [2]
FT	VARIANT	7038	7038	a in BH10, H9; g in BH8 [2], HXB3 [7]
FT	VARIANT	7119	7119	a in BH10; HXB3 [7], H9; g in BH8 [2]
FT	VARIANT	7121	7123	cca in BH10, H9; cac in BH8 [2],
FT				HXB3 [7]
FT	VARIANT	7171	7172	gt in BH10, H9; aa in BH8 [2], HXB3 [7]
FT	VARIANT	7187	7187	a in BH10, H9; g in BH8 [2], HXB3 [7]
FT	VARIANT	7272	7273	aa in BH10, H9; gc in BH8 [2], HXB3 [7]
FT	VARIANT	7291	7291	a in BH10, BH8 [2], H9; c in HXB3 [7]
FT	VARIANT	7343	7343	g in BH10, BH8 [2]; a in HXB3 [7], H9
FT	VARIANT	7439	7454	gtttaatagtacttgg in BH10, HXB3 [7],
FT				and H9
FT	VARIANT	7461	7461	a in BH10, BH8 [2]; g in HXB3 [7], H9
FT	VARIANT	7491	7493	c in BH10, BH8 [2]; a in HXB3 [7], H9
FT	VARIANT	7521	7521	a in BH10, BH8 [2]; t in HXB3 [7], H9
FT	VARIANT	7574	7574	t in BH10, BH8 [2]; c in HXB3 [7], H9
FT	VARIANT	7636	7636	g in BH10, BH8 [2]; a in HXB3 [7], H9
FT	VARIANT	7636	7637	cg in BH10, HXB3 [7], H9; gc in BH8 [2]
FT	VARIANT	7645	7645	a in BH10, BH8 [2], H9; g in HXB3 [7]
FT	VARIANT	8060	8061	ca in BH10, BH8 [2], H9; ac in H
FT	VARIANT	8127	8127	a in BH10, BH8 [2], H9; c in HXB [7]
FT	VARIANT	8131	8131	t in BH10, BH8 [2], H9; c in HXB3 [7]
FT	VARIANT	8135	8135	c in BH10, BH8 [2], H9; g in HXB3 [7]
FT	VARIANT	8257	8257	g in BH10, BH8, HXB3; a in H9
FT	VARIANT	8273	8273	t in BH10, BH8, HXB3; g in H9
FT	VARIANT	8364	8364	g in BH10, HXB3 [7]; a in BH8 [2], H9
FT	CDS	8403	8454	TAT protein, exon 3 (AA at 8410)
FT	CDS	8409	8623	ART protein, exon 3 (putative; AA at 8411)
FT				t in BH10, HXB3 [7], clone 12 cDNA
FT				[21]; a in BH8 [2]; c in H9
FT	VARIANT	8464	8464	g in BH10, BH8, HXB3, clone 12 cDNA
FT				[21]; a in H9

FT	VARIANT	8657	8657	g in BH10, BH8 [2]; a in HXB3 [7], H9, clone 12 cDNA [21]
FT				
FT	VARIANT	8672	8672	g in BH10, HXB3 [7], clone 12 cDNA [21], H9; a in BH8 [2]
FT				
FT	VARIANT	8682	8682	g in BH10, HXB3 [7], clone 12 cDNA [21], H9; a in BH8 [2]
FT				
FT	VARIANT	8748	8748	g in BH10, HXB3 [7], clone 12 cDNA [21], H9; t in BH8 [2]
FT				
FT	VARIANT	8758	8758	g in BH10, H9; c in BH8 [2]; a in HXB3 [7], clone 12 cDNA [21]
FT				
FT	VARIANT	8771	8771	t in BH10, HXB3 [7], clone 12 cDNA [21], H9; c in BH8 [2]
FT				
FT	CDS	8827	9447	27K protein, exon 3 (first expressed exon)
FT				
FT	VARIANT	8857	8857	g in BH10, BH8, HXB3, clone 12 cDNA [21]; a in H9
FT				
FT	VARIANT	8924	8924	c in BH10, HXB3 [7], clone 12 cDNA [21], H9; t in BH8 [2]
FT				
FT	VARIANT	8967	8967	c in BH10, clone 12 cDNA [21], H9; t in BH8 [2]
FT				
FT	VARIANT	8978	8978	a in BH10, clone 12 cDNA [21], H9; c in BH8 [2]
FT				
FT	VARIANT	8985	8985	t in BH10, clone 12 cDNA [21], H9; c in BH8 [2]
FT				
FT	VARIANT	8997	8997	a in BH10, BH8; c in H9, clone 12 cDNA [21]
FT				
FT	VARIANT	8994	8994	c in BH10, clone 12 cDNA [21], H9; t in BH8 [2]
FT				
FT	VARIANT	9018	9018	g in BH10, BH8; a in H9, clone 12 cDNA [21]
FT				
FT	RPT	9116	9748	3' LTR
FT				
FT	VARIANT	9168	9168	t in BH10, clone 12 cDNA [21]; c in BH8 [2]
FT				
FT	VARIANT	9187	9187	g in BH8 [2], H9, clone 12 cDNA [21]; a in BH10 [2]
FT				
FT	VARIANT	9216	9216	g in BH10, BH8; a in H9, clone 12 cDNA [21]
FT				
FT	VARIANT	9222	9223	ga in BH10, clone 12 cDNA [21], H9; ag in BH8 [2]
FT				
FT	VARIANT	9278	9279	g in BH10, BH8, clone 12 cDNA [21]; t in H9
FT				
FT	VARIANT	9283	9283	t in BH10, BH8, clone 12 cDNA [21]; g in H9
FT				
FT	VARIANT	9284	9284	t in BH10, H9, clone 12 cDNA [21]; a in BH8 [2]
FT				
FT	VARIANT	9291	9291	a in BH10, BH8, clone 12 cDNA [21]; g in H9
FT				
FT	VARIANT	9297	9297	c in BH10, clone 12 cDNA [21], H9; t in BH8 [2]
FT				
FT	VARIANT	9354	9354	g in BH10, HIVDSM, H9; t in BH8 [2]
FT				
FT	VARIANT	9408	9408	a in BH10, BH8; g in H9, clone 12 cDNA [21]
FT				
FT	VARIANT	9448	9448	c in BH10; t in BH8 [2], H9, clone 12 cDNA
FT				
FT	VARIANT	9536	9563	c in BH10, BH8, clone 12 cDNA [21]; g in H9
FT				
FT	RPT	9570	9556	R repeat 3' copy
FT				
FT	VARIANT	9616	9616	g in HXB2; a in H9, clone 12 cDNA [21]
FT				
FT	VARIANT	9621	9621	g in HXB2; a in H9, clone 12 cDNA [21]
FT				
FT	VARIANT	9663	9663	t in BH10, H9; tg in clone 12 cDNA [21]
FT				
FT	POLYA	9666	9666	TAT, ART, 27K mRNA exon 3 end (poly-A site) [10],[18],[21]
FT				
FT	POLYA	9666	9666	genomic mRNA end (poly-A site) [10]

Initial Score	=	1872	Optimized Score	=	2186	Significance	=	0.00
Residue Identity	=	88%	Matches	=	2255	Mismatches	=	157
Gaps	=	110	Conservative Substitutions	=			=	0

[illegible]

ATCCCTTTEAGCCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTTGCATTCTAAAATGTAATAATAAGAC
 6680 6690 6700 6710 6720 6730 6740
 GTTCAATGGAACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATC
 6950 6960 6970 6980 6990 7000 7010 7020
 AACTCAACTGCTGTTCGAATGGCAGTCTAGCAGAAGAAGAGGTTAGTAATTAGATCTGCCAATTTTCACAGACAA
 7030 7040 7050 7060 7070 7080 7090
 TGCTAAAACCATTAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAG
 7100 7110 7120 7130 7140 7150 7160
 AAAAAGTATCCGTATCCAGAGGGGACCCAGGGAGAGCATTTGTTACAATAGGAAAAATAGGAAATATGAGACA
 7170 7180 7190 7200 7210 7220 7230
 AGCACATTGTACATTAGTAGAGCAAAATGCAATGCCACTTTAAACAGATAGCTAGCAATTAAGAGAACA
 7240 7250 7260 7270 7280 7290 7300
 ATTTGGAAATATTAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTT
 7310 7320 7330 7340 7350 7360 7370 7380
 TAAATGTGGAGGGGAATTTTCTACTGTAAATCAACACAACCTGTTTAAATAGTACTTGGTTTAAATAGTACTTG
 7390 7400 7410 7420 7430 7440 7450
 GAGTACTGAAGGCTCAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
 7460 7470 7480 7490 7500 7510 7520
 AAACATGTGGCAGGAAGTAGGAAGAGCAATGTATGCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAAA
 7530 7540 7550 7560 7570 7580 7590
 TATTACAGGGCTGCTATTAAACAGAGATGGTGGTAATAACAACAATGGGTCCGAGATCTTCAGACCTGGAGG
 7600 7610 7620 7630 7640 7650 7660
 1460 1470 1480 1490 1500 1510 1520

AGGCGATATGAGGGACATTTGAGGAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGT
|||||
AGGCGATATGAGGGACATTTGAGGAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGT
7670 7680 7690 7700 7710 7720 7730 7740

1530 1540 1550 1560 1570 1580 1590
AGCAGCCACCAAGGCAGAGAAAGAGTGGTGTCAGAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCT
|||||
AGCAGCCACCAAGGCAGAGAAAGAGTGGTGTCAGAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCT
7750 7760 7770 7780 7790 7800 7810

1600 1610 1620 1630 1640 1650 1660
TGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCACGGTCAATGACGCTGACGGTACAGGCCAGACAATT
|||||
TGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCACGGTCAATGACGCTGACGGTACAGGCCAGACAATT
7820 7830 7840 7850 7860 7870 7880

1670 1680 1690 1700 1710 1720 1730
ATTGTTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACT
|||||
ATTGTTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACT
7890 7900 7910 7920 7930 7940 7950

1740 1750 1760 1770 1780 1790 1800 1810
CACAGTCTAGGTCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAGGATCAACAGCT
|||||
CACAGTCTAGGTCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAGGATCAACAGCT
7960 7970 7980 7990 8000 8010 8020

1820 1830 1840 1850 1860 1870 1880
CCTGAGGAATTTGGGTTGCTCTGGAAGAACTCATTTGCACCACTGCTGTGCCTTGAATGCTAGTTGGAGTAA
|||||
CCTGAGGAATTTGGGTTGCTCTGGAAGAACTCATTTGCACCACTGCTGTGCCTTGAATGCTAGTTGGAGTAA
8030 8040 8050 8060 8070 8080 8090 8100

1890 1900 1910 1920 1930 1940 1950
TAAATCTCTGGAGAACATTTGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAG
|||||
TAAATCTCTGGAGAACATTTGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAG
8110 8120 8130 8140 8150 8160 8170

1960 1970 1980 1990 2000 2010 2020
CTTAATACACTCCTTAATTTGAAGAAATCGCAAACCCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGA
|||||
CTTAATACACTCCTTAATTTGAAGAAATCGCAAACCCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGA
8180 8190 8200 8210 8220 8230 8240

2030 2040 2050 2060 2070 2080 2090
TAAATGGGCACATTTTGTGGAATTTGTTTAACATAACAAATTGGCTGTGGTATATAAAAATTATTCATAATGAT
|||||
TAAATGGGCACATTTTGTGGAATTTGTTTAACATAACAAATTGGCTGTGGTATATAAAAATTATTCATAATGAT
8250 8260 8270 8280 8290 8300 8310

2100 2110 2120 2130 2140 2150 2160 2170
AGTAGGABGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATA
|||||
AGTAGGABGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATA
8320 8330 8340 8350 8360 8370 8380

2180 2190 2200 2210 2220 2230 2240
TTCACCATTTATGTTTCAGAGCCCACCTCCCCAACCCGAGGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGA
|||||
TTCACCATTTATGTTTCAGAGCCCACCTCCCCAACCCGAGGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGA
8390 8400 8410 8420 8430 8440 8450 8460

8470 8480 8490 8500 8510 8520 8530 8540

```

AGGTGGAGAGAGAGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCT
|||||
AGGTGGAGAGAGAGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCT
      8470      8480      8490      8500      8510      8520      8530

      2320      2330      2340      2350      2360      2370      2380
GCCGAGCCCTTGTGCTTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAAC
|||||
GCCGAGCCCTTGTGCTTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAAC
      8540      8550      8560      8570      8580      8590      8600

      2350      2400      2410      2420      2430      2440      2450
TTCTGGGACGCGAGGGGTGGGAAGCCCTCAAATATTGGTGGGAATCTCCTACAGTATTGGAGTCAGGAACTAA
|||||
TTCTGGGACGCGAGGGGTGGGAAGCCCTCAAATATTGGTGGGAATCTCCTACAGTATTGGAGTCAGGAGCTAA
      8610      8620      8630      8640      8650      8660      8670

2460
AG
||
AG
X

```

3. KUNZ-158-CL32. S68

HIVMAL06 Human lymphadenopathy virus (MAL isolate), complet

ID HIVMAL06 standard; RNA; 9229 BP.

XX

AC X04415.

XX

DT 17-OCT-1986 (incorporated)

XX

DE Human lymphadenopathy virus (MAL isolate), complete genome.

XX

KW acquired immune deficiency syndrome; env gene; gag gene; genome;

KW long terminal repeat; pol gene; polyprotein; provirus;

KW reverse transcriptase.

XX

DS Human lymphadenopathy virus

OC Viridae; ss-RNA enveloped viruses; Retroviridae.

XX

RN [1] (bases 1-9229)

RA Alizon M., Vain-Hobson S., Montagnier L., Sonigo P.;

RT "Genetic variability of the AIDS virus: Nucleotide sequence

RT analysis of two isolates from African patients";

RL Cell 46:63-74(1986).

XX

CC Acquired immune deficiency syndrome (AIDS) is caused by a
 CC retrovirus known by several different names, probably representing
 CC two separate strains: human T-cell lymphotropic virus-III
 CC (HTLV-III) and lymphadenopathy-associated virus (LAV) are thought
 CC to be one strain; and AIDS-associated retrovirus type 2 (ARV-2) the
 CC other. All three viruses, whose sequences do not differ by more
 CC than about 6% are believed to belong to the retroviral subfamily
 CC Lentiviridae, or "slow" viruses. For the details of the annotation
 CC and for other pertinent references, see the HIV reference entry.

XX

FH	Key	From	To	Description
----	-----	------	----	-------------

FH

FT	RPT	1	93	R repeat 5' copy
----	-----	---	----	------------------

FT	RPT	1	177	5' LTR
----	-----	---	-----	--------

FT	SYDE	179	186	primer (Lys-tRNA) binding site
----	------	-----	-----	--------------------------------

FT	CDS	350	1337	gag polyprotein
----	-----	-----	------	-----------------

FT	CDS	1563	4671	pol polyprotein (NH2-terminus
----	-----	------	------	-------------------------------

FT				uncertain; EA at 1963)
----	--	--	--	------------------------

FT	CDR	4515	5154	sor 23K protein
FT	CDR	5134	5424	urfc
FT	CDR	5405	5515	tat protein, exon 2 (first expressed
FT				exon)
FT	CDR	5755	8373	envelope polyprotein precursor
FT	CDR	7855	8007	tat protein, exon 3 (AA at 7960)
FT	CDR	8350	9005	27K protein
FT	RPT	8675	9225	3' LTR
FT	RPT	5134	5225	R repeat 3' copy

XX
 SB Sequence 5225 BP: 3355 A; 1627 C; 2204 G; 2043 T; 0 other;

Initial Score	=	916	Optimized Score	=	2060	Significance	=	0.00
Residue Identity	=	94%	Matches	=	2085	Mismatches	=	350
Gaps	=	43	Conservative Substitutions	=			=	0

```

X      10      20      30      40      50      60      70
AAGAGCAGAGAGAGAGTGGCAATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAAATGGG
|||||
AAGAGCAGAGAGAGTGGCAATGAGAGTGAAGGAGAGATACA-GAGGAATTATCAAAA--CTGGTGGAGATGGG
5780      5790      5800      5810      5820      5830      5840

      80      90      100      110      120      130      140
GCACCATGCTCCCTGGGATATTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGG
|||
GCATGATGCTCCCTGGGATATTGATGACCTGTAGTATTGCAGAGATTTGTGGGTTACAGTTTATTATGGGG
5850      5860      5870      5880      5890      5900      5910

      150      160      170      180      190      200      210
TACCTGTGTGGAGAGAGCAGCAGCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTAC
|||||
TACCTGTGTGGAGAGAGCAGCAGCACTCTATTTTGTGCATCAGATGCTAAATCATATGAAACAGAGGTAC
5920      5930      5940      5950      5960      5970      5980      5990

      220      230      240      250      260      270      280
ATAATGTGTTTGGGCGACACATGCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGA
|||||
ATAACATCTGGGCTACACATGCTGTGTACCCACGGACCCCAACCCACAAGAAATAGAACTGGAAAATGTCA
6000      6010      6020      6030      6040      6050      6060

      290      300      310      320      330      340      350      360
CAGGAAATTTTAAACATGTGGGAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATC
|||||
CAGGAGGGTTTAAACATGTGGGAAAATAACATGGTGGAGCAGATGCATGAGGATATAATCAGTTTATGGGATC
6070      6080      6090      6100      6110      6120      6130

      370      380      390      400      410      420
AAACCCCTAAAGCAGTGTGTAAATTAACCCCACTCTGTGTTAGTTTAAAGTGCAGTGTATTG---GGGAATG
|||||
AAACCCCTAAAGCAGTGTGTAAAGTAAACCCCACTCTGTGTCAGTTTAAAGTGCAGTGTATTGGAATGGGACTG
6140      6150      6160      6170      6180      6190      6200

      430      440      450      460      470      480      490
CTATTAAT---ACCAATACTAGTAATACCAATAGTAGTAGCGGGGAAATGATGATGGAGAAAGGAGAGATAA
|||
CTGTGAATGGGACTAATGCTGGGACT---AATAGGACTAATGCAGAAATTGAAAATGGAAATTGGAGAAAGTGA
6210      6220      6230      6240      6250      6260      6270

      500      510      520      530      540      550      560      570
AAACCTGCTCTTTCAATATCAGCACAAGNATAAGAGGTAAGGTGCAGAAAGAAATATGCATTTTTTTATAAAC
|||||
AAACCTGCTCTTTCAATATAACCCAGTAGGAAGTGATAAAAGGC---AAGAATATGCAACTTTTTTATAACC
6280      6290      6300      6310      6320      6330      6340

      580      590      600      610      620      630
TTGCTATTAATACCAATA---GATGATGAAGTACCACTATACCTTGACAAGTTGTAACACCTCAGTCATTA

```


1111 11 11 111 1111 1111 1111 11 11 1111 1111 1111 1111 1111 1111 1111 1111
 TTGATCTATGACAAATAGATGATGATGATAATAGTAGTTATAGGCTAATAAATTGTAATACCTCAGTAATTA
 6360 6360 6370 6380 6390 6400 6410
 640 650 660 670 680 690 700 710
 CAGGAGCCTGTCGCAAGGTATUCCTTGAGCCAATTCCCACATTATTGTGCCCCGGCTGGTTTTGCGATTG
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 CAGGAGCCTGTCGCAAGGTAACTTTGATCCAATTCCCACATTATTGTGCCCCAGCTGGTTTTGCAATTG
 6420 6430 6440 6450 6460 6470 6480
 720 730 740 750 760 770 780
 TAAATGTAAATAAAGACCTTCAATGGAACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATG
 111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 TAAATGTAAATAAAGAGTTCAATGGAACAGGAAATATGTAAATGTCAGTACAGTACAATGTACACATG
 6490 6500 6510 6520 6530 6540 6550 6560
 750 800 810 820 830 840 850
 GAATTAAGGACCTAGTATCAACTCAACTGCTGTTGAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGAT
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 GAATTAAGGACCTAGTATCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGATAATTGATTAGAT
 6570 6580 6590 6600 6610 6620 6630
 860 870 880 890 900 910 920
 CTGCGAATTTCAAGACCAATGCTAAAACCATATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAA
 111 111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 CTGTAATCTCAAGACCAATGCTAAAACCATATAGTACAGCTTAATGAACTGTAAATTAATTGTACAA
 6640 6650 6660 6670 6680 6690 6700
 930 940 950 960 970 980 990
 GAGCCAAACCAATATAGCAAGAAAGTATCCGTATCCAGAGGGGACCAAGGGAGAGCATTTGTTACAATAGGAA
 111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 GAGCCAAACCAATATAGCAAGAAAGTATCCGTATCCAGAGGGGACCAAGGGAGAGCATTTGTTACAATAGGAA
 6710 6720 6730 6740 6750 6760 6770
 1000 1010 1020 1030 1040 1050 1060 1070
 AAATAGGAAATATAGCAAGAAAGTATCCGTATCCAGAGGGGACCAAGGGAGAGCATTTGTTACAATAGGAA
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 TAAATAGGAAATATAGCAAGAAAGTATCCGTATCCAGAGGGGACCAAGGGAGAGCATTTGTTACAATAGGAA
 6780 6790 6800 6810 6820 6830 6840
 1080 1090 1100 1110 1120 1130 1140
 CTATCAAAATTAAGCAAGCAATTTGGAATTAATAAACAATAATCTTTAAGCAATCTCAGGAGGGGACCCAG
 111
 CTATCAAAATTAAGCAAGCAATTTGGAATTAATAAACAATAATCTTTAAGCAATCTCAGGAGGGGACCCAG
 6850 6860 6870 6880 6890 6900 6910
 1150 1160 1170 1180 1190 1200 1210
 AAATTTGTAAGCAAGCAATTTGGAATTAATAAACAATAATCTTTAAGCAATCTCAGGAGGGGACCCAG
 1111
 AAATTTGTAAGCAAGCAATTTGGAATTAATAAACAATAATCTTTAAGCAATCTCAGGAGGGGACCCAG
 6920 6930 6940 6950 6960 6970 6980
 1220 1230 1240 1250 1260 1270 1280
 CTCTGTTTAAATTAAGCAAGCAATTTGGAATTAATAAACAATAATCTTTAAGCAATCTCAGGAGGGGACCCAG
 111
 CTCTGTTTAAATTAAGCAAGCAATTTGGAATTAATAAACAATAATCTTTAAGCAATCTCAGGAGGGGACCCAG
 6990 7000 7010 7020 7030 7040 7050
 1290 1300 1310 1320 1330 1340 1350
 GCAATTAATTAAGCAAGCAATTTGGAATTAATAAACAATAATCTTTAAGCAATCTCAGGAGGGGACCCAG
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 GCAATTAATTAAGCAAGCAATTTGGAATTAATAAACAATAATCTTTAAGCAATCTCAGGAGGGGACCCAG
 7060 7070 7080 7090 7100 7110 7120
 1360 1370 1380 1390 1400 1410 1420
 GCAATTAATTAAGCAAGCAATTTGGAATTAATAAACAATAATCTTTAAGCAATCTCAGGAGGGGACCCAG
 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 GCAATTAATTAAGCAAGCAATTTGGAATTAATAAACAATAATCTTTAAGCAATCTCAGGAGGGGACCCAG

[illegible]

FT	517	1	134	5' LTR
FT	517E	132	134	primer (lys-tRNA) binding site
FT	518	135	1838	gag polyprotein
FT	518	1839	4342	pol polyprotein (NH2-terminus
FT				uncertain: AA at 1904)
FT	519	4343	5165	env 23K protein
FT	519	5166	5325	orfC
FT	520	5326	5550	orfD
FT	520	5551	5731	envelope polyprotein precursor
FT	520	5732	5847	27K protein
FT	521	5848	6175	3' LTR
FT	521	6176	6178	R repeat 3' copy

XX
 SQ Sequence 5172 BPV 5033 A: 1832 C: 2179 G: 2032 T: 0 other;

Initial Score = 1246 Optimized Score = 1904 Significance = 0.00
 Residue Identity = 70% Matches = 1971 Mismatches = 443
 Gaps = 105 Conservative Substitutions = 0

X	10	20	30	40	50	60
AMEYDORSA--DPAQNEFGUCKATGAGAGTGAAGGAGAAATATCAG---- <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>						
QPSAGGAGNXX--LUCGATTTACTTGATAGAATAACAGAAAGAGCAGAAGACAGTGGCAATGAGAGCGAGGGG						
X	5720	5730	5740	5750	5760	5770
	70	80	90	100	110	120
RPQVQDPAATGACCTGCTC--CTTG--GGATATTG--ATGAT-CT-GTAGTGCTACAGAAAAATTGT-GG						
GATGAGAGGAG--ATGCTCAGAAATGGTGGAAATGGGGCATCATGCTCCTTGGGATATTGATGACCTGTAGT						
X	5750	5800	5810	5820	5830	5840
	130	140	150	160	170	180
GTGAGAGTCTATATAGGCTACCTGTGTGGAAGGAAGCAACCACCACTCTATTTTG-TGCATCAGATGCTAA						
GCTACAGACCATCTGTGGGTACAGTTTATTATGGGG----TGCCTGTATGGAAGGAAGCAACCACCACTCT						
X	5850	5900	5890	5880	5900	5910
	200	210	220	230	240	250
AGGATATGATGATGCTGCTACCTGCTATG--TTGGGCCACA-CATGCCTGTGTACCCACAGACC--CCAAC						
ATTGATGATGATGCTGCTACCTGCTATG--TTGGGCCACA-CATGCCTGTGTACCCACAGACC--CCAAC						
X	5930	5940	5950	5960	5970	5980
	270	280	290	300	310	320
CCAG--GAGAGCTGATGCTGCTACCTGCTATG--TTGGGCCACA-CATGCCTGTGTACCCACAGACC--CCAAC						
CCAG--GAGAGCTGATGCTGCTACCTGCTATG--TTGGGCCACA-CATGCCTGTGTACCCACAGACC--CCAAC						
X	6000	6010	6020	6030	6040	6050
	330	340	350	360	370	380
ATC--ATGATGCTGCTACCTGCTATG--TTGGGCCACA-CATGCCTGTGTACCCACAGACC--CCAAC						
AT--AAGATGCTGCTACCTGCTATG--TTGGGCCACA-CATGCCTGTGTACCCACAGACC--CCAAC						
X	6060	6070	6080	6090	6100	6110
	420	430	440	450	460	
GTGATGCTGCTACCTGCTATG--TTGGGCCACA-CATGCCTGTGTACCCACAGACC--CCAAC						
GTGATGCTGCTACCTGCTATG--TTGGGCCACA-CATGCCTGTGTACCCACAGACC--CCAAC						
X	6130	6140	6150	6160	6170	6180
	470	480	490	500	510	520
GGG--GAGAGCTGATGCTGCTACCTGCTATG--TTGGGCCACA-CATGCCTGTGTACCCACAGACC--CCAAC						
GGG--GAGAGCTGATGCTGCTACCTGCTATG--TTGGGCCACA-CATGCCTGTGTACCCACAGACC--CCAAC						
X	6190	6200	6210	6220	6230	6240
	530					
GGG--GAGAGCTGATGCTGCTACCTGCTATG--TTGGGCCACA-CATGCCTGTGTACCCACAGACC--CCAAC						
GGG--GAGAGCTGATGCTGCTACCTGCTATG--TTGGGCCACA-CATGCCTGTGTACCCACAGACC--CCAAC						
X	6250					


```

1110      1120      1130      1140      1150      1160      1170
AGGAGGCTTGATAGAGTTTAAAGAAAGTTTCTTTTGGCTGTACTTTTCTATAGTGAATAGAGTTAGGCAGGGGATATTC
|||||
AGGAGGCTTGATAGAGTTTAAAGAAAGTTTCTTTTGGCTGTACTTTTCTTTAGTAAATAGAGTTAGGCAGGGGATACTC
7235      7240      7250      7260      7270      7280      7290

1180      1190      1200      1210      1220      1230      1240
ACCATTTATCTTTTCAAGTCCCAATCTCCCAACCCCGAGAGGGACCCCGACAGGGCCCGAAGGAATAGAAGAAGAAGG
|||||
ACCATTTATCTTTTCAAGTCCCAATCTCCCAACCCCGAGAGGGACCCCGACAGGGCCCGAAGGAACAGAAGAAGAAGG
7300      7310      7320      7330      7340      7350      7360      7370

1250      1260      1270      1280      1290      1300      1310
TGGAGACGAGAGTCCAGAGACAGTCTGATTCGATTAGTGAACCGGATCCTTAGCACTTATCTGGGACGATCTGCG
|||||
TGGAGACGAGAGTCCAGAGACAGTCTGATTCGATTGCTGAACCGGATTCTCGGCACTTATCTGGGACGACCTGCG
7380      7390      7400      7410      7420      7430      7440

1320      1330      1340      1350      1360      1370      1380
GAGCCTTGATGATCTTCACTTACACCTCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACTTC
|||||
GAGCCTTGATGATCTTCACTTACACCTCTTGAGAGACTTAACTCTTAATTGCAGTGAGGATTGTAGAACTTC
8050      8060      8070      8080      8090      8100      8110

1390      1400      1410      1420      1430      1440      1450
TGGAGACGAGAGTCCAGAGACAGTCTGATTCGATTAGTGAACCGGATCCTTAGCACTTATCTGGGACGATCTGCG
|||||
TGGAGACGAGAGTCCAGAGACAGTCTGATTCGATTGCTGAACCGGATTCTCGGCACTTATCTGGGACGACCTGAGG
8120      8130      8140      8150      8160      8170      8180      X

```


1470
1540
1610
880
950
1020
1090
1140
1210

600
610
620
630
640
650
660
670
680
690
700
710
720
730
740
750
760
770
780
790
800
810
820
830
840
850
860
870
880
890
900
910
920
930
940
950
960
970
980
990
1000
1010
1020
1030
1040
1050
1060
1070
1080
1090
1100
1110
1120
1130
1140
1150
1160
1170
1180
1190
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
1300
1310
1320
1330
1340
1350
1360
1370
1380
1390
1400
1410
1420
1430
1440
1450
1460
1470
1480
1490
1500
1510
1520
1530
1540
1550
1560
1570
1580
1590
1600
1610
1620
1630
1640
1650
1660
1670
1680
1690
1700
1710
1720
1730
1740
1750
1760
1770
1780
1790
1800
1810
1820
1830
1840
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030
2040
2050
2060
2070
2080
2090
2100
2110
2120
2130
2140
2150
2160
2170
2180
2190
2200
2210
2220
2230
2240
2250
2260
2270
2280
2290
2300
2310
2320
2330
2340
2350
2360
2370
2380
2390
2400
2410
2420
2430
2440
2450
2460
2470
2480
2490
2500
2510
2520
2530
2540
2550
2560
2570
2580
2590
2600
2610
2620
2630
2640
2650
2660
2670
2680
2690
2700
2710
2720
2730
2740
2750
2760
2770
2780
2790
2800
2810
2820
2830
2840
2850
2860
2870
2880
2890
2900
2910
2920
2930
2940
2950
2960
2970
2980
2990
3000
3010
3020
3030
3040
3050
3060
3070
3080
3090
3100
3110
3120
3130
3140
3150
3160
3170
3180
3190
3200
3210
3220
3230
3240
3250
3260
3270
3280
3290
3300
3310
3320
3330
3340
3350
3360
3370
3380
3390
3400
3410
3420
3430
3440
3450
3460
3470
3480
3490
3500
3510
3520
3530
3540
3550
3560
3570
3580
3590
3600
3610
3620
3630
3640
3650
3660
3670
3680
3690
3700
3710
3720
3730
3740
3750
3760
3770
3780
3790
3800
3810
3820
3830
3840
3850
3860
3870
3880
3890
3900
3910
3920
3930
3940
3950
3960
3970
3980
3990
4000
4010
4020
4030
4040
4050
4060
4070
4080
4090
4100
4110
4120
4130
4140
4150
4160
4170
4180
4190
4200
4210
4220
4230
4240
4250
4260
4270
4280
4290
4300
4310
4320
4330
4340
4350
4360
4370
4380
4390
4400
4410
4420
4430
4440
4450
4460
4470
4480
4490
4500
4510
4520
4530
4540
4550
4560
4570
4580
4590
4600
4610
4620
4630
4640
4650
4660
4670
4680
4690
4700
4710
4720
4730
4740
4750
4760
4770
4780
4790
4800
4810
4820
4830
4840
4850
4860
4870
4880
4890
4900
4910
4920
4930
4940
4950
4960
4970
4980
4990
5000
5010
5020
5030
5040
5050
5060
5070
5080
5090
5100
5110
5120
5130
5140
5150
5160
5170
5180
5190
5200
5210
5220
5230
5240
5250
5260
5270
5280
5290
5300
5310
5320
5330
5340
5350
5360
5370
5380
5390
5400
5410
5420
5430
5440
5450
5460
5470
5480
5490
5500
5510
5520
5530
5540
5550
5560
5570
5580
5590
5600
5610
5620
5630
5640
5650
5660
5670
5680
5690
5700
5710
5720
5730
5740
5750
5760
5770
5780
5790
5800
5810
5820
5830
5840
5850
5860
5870
5880
5890
5900
5910
5920
5930
5940
5950
5960
5970
5980
5990
6000
6010
6020
6030
6040
6050
6060
6070
6080
6090
6100
6110
6120
6130
6140
6150
6160
6170
6180
6190
6200
6210
6220
6230
6240
6250
6260
6270
6280
6290
6300
6310
6320
6330
6340
6350
6360
6370
6380
6390
6400
6410
6420
6430
6440
6450
6460
6470
6480
6490
6500
6510
6520
6530
6540
6550
6560
6570
6580
6590
6600
6610
6620
6630
6640
6650
6660
6670
6680
6690
6700
6710
6720
6730
6740
6750
6760
6770
6780
6790
6800
6810
6820
6830
6840
6850
6860
6870
6880
6890
6900
6910
6920
6930
6940
6950
6960
6970
6980
6990
7000
7010
7020
7030
7040
7050
7060
7070
7080
7090
7100
7110
7120
7130
7140
7150
7160
7170
7180
7190
7200
7210
7220
7230
7240
7250
7260
7270
7280
7290
7300
7310
7320
7330
7340
7350
7360
7370
7380
7390

128

1320 1330 1340

1350 1360 1370 1380 1390 1400 1410 1420

1430 1440 1450 1460 1470 1480 1490

1500 1510 1520 1530 1540 1550

2350 2360 2370 2380 2390 2400 2410

2420 2430 2440 2450 2460 2470 2480

2490 2500 2510 2520 2530 2540 2550

1700 1710 1720 1730 1740 1750 1760 1770

2560 2570 2580 2590 2600 2610 2620

2630 2640 2650 2660 2670 2680 2690 2700

2710 2720 2730 2740 2750 2760

2770 2780 2790 2800 2810 2820

2830 2840 2850 2860 2870 2880 2890 2900

RL
XX
RN
RA
RA
RT
RT
RL
XX
CC
XX
FH
FH
FT
FT
FT
FT
FT
FT
FT
FT
FT
FT
FT
FT
F
F
F
F
F
F
F
F
F

[illegible][illegible][illegible][illegible][illegible][illegible][illegible][illegible][illegible]

RL
XX
RN
RA
RA
RT
RT
RL
XX
CC
XX
FH
FH
FT
FT
FT
FT
FT
FT
FT
FT
FT
FT
FT
FT
F
F
F
F
F
F
F
F
F

[illegible][illegible][illegible]

1 111 11111 1 111 1111 1 11 1 11 11111 1 1 11 111 1
TTASTGTTACAGAAAGCTTTTATCTCTGGGAGA---AT-ACAGT-----CACAGAAC---AGGCAATAGAGG
5940 5950 5960 5970 5980 5990

200 250 300 310 320 330
TASATATTGATTAATCTGAGAGGAAATTTTAACATGTGGAAAA--ATGA-CAT--GGTAGAAC----AGAT
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
ACCATATAGCAAGCTCTTCTGAGAGCTCAATTAAGGCTTGTGTAAAATTATCCCCATTATGCATTACTATGAGAT
6000 6010 6020 6030 6040 6050 6060

340 350 360 370 380 390
GCATGAGGATATATATGAG-TTTAT-AGGAT---CA-AAGCCTAAA-GCCATGTGTAAAATTAACCCC-AC-T
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
GCCATTAAGCTGAGGAGGATGAGTGGGATGACAAAATCATCAACAACAATAACAACAGCAGCACCAACAT
6070 6080 6090 6100 6110 6120 6130

400 410 420 430 440 450 460
CTCTGTTAGT-TTAAGATGCACTGCTTTGGGGAATGCTACTAATACCAATACTAG-TAATACCAATAGTAGT
11 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
CAGCATCAATATCAGAAAGAAATAGACATGCTCAATGAGACTAGTTCTTGTA-TAGCTCAGAATAATTGCACA
6140 6150 6160 6170 6180 6190 6200

470 480 490 500 510 520 530
ATCTTGGGAT-ATTA-TGATGAGAGAGGAGATAAAAACTGCTCTTTCAATATCAGCACAAGNATAAGAG
11 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
GCTTTGGGAGGAGAGCAATGATAGCTGTAAATTCACCATGACAGGGTTAAAAAGAG-ACAAGACAAGGA
6210 6220 6230 6240 6250 6260 6270

540 550 560 570 580 590 600
GTAAAGTGTGAGAG--AATATGC--ATTT-TTTTATAAACTTGATATAATACCAATAGATAATGATACTAC
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
GTACATGAGAGCTGGTACTGTACAGATTTGGTTTGTGAAC--AAGGGAATAGCACT-GATAATGAAAGCAG
6280 6290 6300 6310 6320 6330 6340

610 620 630 640 650 660 670
CAGCTATAGCTTGACAGTTGTAAACCTCAGTCATTACACAGGCCTGT-CCAAAGGTATCCTTTGAGCCAA
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
ATGCTACA---TGAACTACTGTAACTTCTGTTATCCAAGAGTCTTGAGACAAACAT-TATTGGGATACTA
6350 6360 6370 6380 6390 6400 6410

680 690 700 710 720 730 740
TTCCCATACATATTGTGCCCCGGCTGGTTTTTGGGATTCTAAATGTAATAATA-AGACGTTCAATGGAACA
11 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
TTAGATTTAGATATTGTGACCTCCAGTTATGCTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTA
6420 6430 6440 6450 6460 6470 6480

750 760 770 780 790 800
GGA--CATGTACAAATGT-CAGCAGCTACAATGTACACATGGA--AT-TAGGC-CAGTAGTATCAACTCA
11 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
TGGCTAAATGTTCTAAGGTGTGCTCTCTTC-ATG--CACAGGATGATGGAGACACAG-ACT-TCTACTTG
6490 6500 6510 6520 6530 6540 6550

810 820 830 840 850 860 870
ACTCTGTGTAATGGCAGTCTAGCAGAAAG-AGAGGTAGTAATTAGATCTGCCAATTTACAGACAATGCTA
11 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
GTTTGGCTTTAATGAACTAGAGCAGAAAATAGAAGTTATATTTA---CTGGC-ATGGTAGGGATAA---TA
6560 6570 6580 6590 6600 6610 6620

880 890 900 910 920 930 940
AAGCATAT-TAG--TACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAA
11 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
GAGCTATATTTAATTTAATTAAGTATTATAATCTAACAATGAATGTAGAAGA-CCAGGA--AAT--AAG--
6630 6640 6650 6660 6670 6680

950 960 970 980 990 1000 1010
AAGATATCTATACCAAGAGGAGAG--GAGAGGAT--TGTTCAT--AGGAGGAAATAGGAAATATGAG
11 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

ACAGTCT---TA-CCA---GTCACCATTATGTCTGGATTGGTTTTCCACTCACAACCACTCACTGATAGGCC
6590 6700 6710 6720 6730 6740 6750

1020 1030 1040 1050 1060 1070 1080
ACAGCA---CATCTAACAATTAG---TAGAGCAAATGCAATGCCACTTTAAAACAGAT--AGCTAGCAAATT--
O-AAGCAGGCATGGT----GTTGGTTTGAGGAAATGGAAGGATGCAATAAAGAGGTGAAACAGACCATTG
6760 6770 6780 6790 6800 6810

1090 1100 1110 1120 1130 1140
---AAGAGAACA---AT-TTGGAAATAATAAAC-AAAT--AATC--TTTAAGCAATCCT-CAGGAGGGGACC
TCATACATCCCAGTATACTGGAACTAACAATACTGATAAAATCAATTTAACGGCTCCTGGAGGAGGAGATC
6820 6830 6840 6850 6860 6870 6880 6890

1150 1160 1170 1180 1190 1200
CAGAAATTSTAACCAACA-GT---TTTAATTGTGGAGGGGAATTTTTCTACTGTAATTCAACACAACACTGTTTA
CGGAA---GTACTTTCATGAGACAAATTGCAGAGGAGAGTTCTCTACTGTAA---AA-TGAATTGGTTT
6900 6910 6920 6930 6940 6950

1210 1220 1230 1240 1250 1260 1270
ATACTACTGGTTTA---ATACTACTTGGAGTACTGAAGGGTCAAATAACACTGAAGGAAGTGACACAATCAC
CTA-AATTGGCTAGAGGATAGGGATGTAATACTACCCAGAGGCCAAAGGA-AC-----GGCATAGAAGGAATTAC
6960 6970 6980 6990 7000 7010 7020

1280 1290 1300 1310 1320 1330 1340
ACTCCCATGCAAGATTAACCAATTTATAAACATGTGGCAGGAAGTAGGAAAAGCAATG---TATGCCCTCCC
-ETGCCGTTCATATTGACAAATAATCAACACTTGGCATAAAGTAGGCAAA--AATGTTTATTTGCCTCCA
7030 7040 7050 7060 7070 7080 7090

1350 1360 1370 1380 1390 1400 1410 1420
ATCAGCGGACAAATTAGATGTTTCATCAAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAACAAC
AFAGAGGGAAGCCTCACGTGTAACTCCACAGTGACCAGTCTCATAGCAAACATAGATTGGACTGATGGAAAC
7100 7110 7120 7130 7140 7150 7160

1430 1440 1450 1460 1470 1480 1490
AATGGGTCCGAG-ATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGA--GAAGTGAATTATATAAA
CA-----AAGTAGATC-----ACCATGA-GTGCAG--AGGTGGCAGAACTGTATCGATTAGAGTTGGGAGAT
7170 7180 7190 7200 7210 7220

1500 1510 1520 1530 1540 1550
TATTAAGTACGAAAAATTGAACCAATTAGAGTAGCACCCACCAAGGCAAGAG-----AAGAGTGGT-GCA--
TATTAATTAGTAGATCACTCCGATTGGCTTGGCCCCACAGATGTGAAGAGGTACACTACTGGTGGCACC
7230 7240 7250 7260 7270 7280 7290

1560 1570 1580 1590 1600 1610 1620
GAGAGAAANPAGNGCAGTGGGKATAGGAGCTTGTTCCTTGGGTTCT-TGGGAGCAGCAGGAAGCACTATGG
TCAGAAATAGAG-AG-GGTTCTTTGTGCTAGGGTCTTGGGTTTTCTCGCAACGGCAGGTTCTGCAATGG
7300 7310 7320 7330 7340 7350 7360

1630 1640 1650 1660 1670 1680 1690
GCGCACGGTCAATGACGCTGACGATACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATT
GCGCGGCGTGGTTCAGGCTGACCGCTCAGTCCGGGACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGC
7370 7380 7390 7400 7410 7420 7430

1700 1710 1720 1730 1740 1750 1760 1770
TGCTGAGCAATGCTGAGGCTGACGCACTGCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCA

TGT TGGGCGCTGGTCAAGAGACACCAAGATTGTTGCGATTGACCGTCTGGGGAACAAAGAACCTCCAGACTA
7440 7450 7460 7470 7480 7490 7500

1780 1790 1800 1810 1820 1830 1840
 GAATCCTGGGTTGTGGYAGGYACCTAAAGGATCAACAGCTCCTGGGGATTGTTGGGGTTCCTCTGGAAAACCTCA
 | | | | | | | | | | | | | | | | |
 GGGTCACTGCCATCGAGAAGTAAGTAGAGGGACCAGGCAGCTGAATGCTTGGGGATGTGCGTTTAGACAAG
 7510 7520 7530 7540 7550 7560 7570 7580

1850 1860 1870 1880 1890 1900 1910
 TTCTGCACCACTGCTGTGCTTGG---AATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATAACA
 TCTGCCACCACTACTGTACCAAGGCCAAATGCAAGT---CTAACACCAGACTGGAACA-AT---GA-TA---
 7550 7600 7610 7620 7630 7640

1920 1930 1940 1950 1960 1970 1980
 TGAUCTGGATGGAGTGGGACAGAGGAATTAAC--AATTACACAAGCTTAATACATTCTTAATTGAAGAATCG
 ---CTTGGCAAGAGTGGGAGCGAAAGGTGACTTCTTGGAGGAAAATATAACA-GCCCTCCTAGAAGAGGGCA
 7650 7660 7670 7680 7690 7700

1990 2000 2010 2020 2030 2040 2050
 CAAACACCAGGCAAGAGAGAATGAACAAGATTATTGGAATTAGATAAATGGGCAAGTTTGTGG-AATTGGTT
 ||| ||||| ||||| | ||||| | || ||| ||||| | ||| ||||| |||||
 CAAATTGCAGCAGGCAAGAACATCTATTAATTACAAAAGTTGAATAGCTGGG-ATGTGTTTGGCAATTGGTT
 7710 7720 7730 7740 7750 7760 7770

2060 2070 2080 2090 2100 2110 2120
 TGAATAAGCAGA CTGGCTGTGGTATATAAAAATATTCA--TAATGATAGTAGGAGGCTTGGTAGGTTTAAAGA
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 TGACCTTGGCTTCCTGGATAAAGTATAT-ACAATATGGAATTTATG-TAGTTGTAGGAGTAATACTGTTAAGA
 7780 7790 7800 7810 7820 7830 7840

2130 2140 2150 2160 2170 2180 2190
 ATAGCTTTTTCGCGACTTTCATAGTGAATAGAGTT-AGGCAGGGATATTCACCA---TTATCGTTTTCAGAC
 ||||| | | | | | | | | | | | | |
 ATAGTGTATCTATATAGTACAAATGCTAGCTA-AGTTAAGGCAGGGGTATAGGCCAGTGTTCTC-TTCCCCAC
 7850 7860 7870 7880 7890 7900 7910

[illegible][illegible]

2320 2330 2340 2350 2360 2370
 CTGGGAGC--CTTG--TGCCTCTTCAGCTACCAACCGCTTGAGAGACTTACTCTTGATTGTA-ACGAGGAT
 ||| || ||| ||| ||| ||| ||| ||| |||
 CTGATACGCCCTTTGACTTGCTATTTCAGCAA-----CTGCAGAACCTTGCTATCGAGAGCATACCA-GAT
 8060 8070 8080 8090 8100 8110 8120

[illegible]

2450 2450
DANIELSON, GENE

7. KUNZ-158-OL39. SEQ
RESIVMXX Simian immunodeficiency virus STLV-III(MAC) provir

ID RESIVMXX standard; DNA; 8646 BP.
XX
AC Y00277;
XX
DT 17-SEP-1987 (annotation)
XX
DE Simian immunodeficiency virus STLV-III(MAC) proviral genome
XX
KW art gene; env gene; envelope protein; gag gene; genome;
KW overlapping genes; pol gene; tat gene.
XX
OS Simian immunodeficiency virus
OC Viridae; ss-RNA enveloped viruses; Retroviridae.
XX
RN [1] (bases 1-8646)
RA Chakrabarti L.
RT ;
RL Submitted (19-AUG-1987) to the EMBL Data Library.
XX
RN [2] (bases 1-8646)
RA Chakrabarti L., Guyader M., Alizon M., Daniel M.D.,
RA Durosiers R.C., Tellois P., Sonigo P.;
RT "Sequence of simian immunodeficiency virus from macaque and its
RT relationship to other human and simian retroviruses";
RL Nature 323:543-547(1987).

XX
CC source: library=infected HUT78 cells; clones=lambda SIV1;

FH	Key	From	To	Description
FT	SITE	1	175	LTR (R)
FT	SITE	1	318	5' LTR
FT	SITE	176	318	LTR (U5)
FT	SITE	486	2068	ORF (gag)
FT	CDS	551	2068	gag gene product (AA 1-506)
FT	CDS	1726	4393	pol gene product (AA 1-1056)
FT	SITE	4793	5467	ORF (Q)
FT	CDS	4826	5467	Q gene product (AA 1-214)
FT	SITE	5255	5633	ORF (X)
FT	CDS	5256	5633	X gene product (AA 1-112)
FT	CDS	5637	5938	R gene product (AA 1-101)
FT	SITE	5758	6083	ORF, exon 1 (tat)
FT	CDS	5788	6083	tat gene product (AA 1-99) (6083 is 2nd base in codon)
FT	SITE	6011	6083	ORF, exon 1 (art)
FT	CDS	6014	6083	art gene product (AA 1-23) (6083 is 1st base in codon)
FT	SITE	6007	6735	ORF (env)
FT	IVS	6084	8300	intron I (tat and art) (8301 is 3rd base in codon)
FT	CDS	6080	6735	envelope protein (AA 1-882)
FT	SITE	8247	8551	ORF, exon 2 (art)
FT	SITE	8251	8394	ORF, exon 2 (tat)
FT	SITE	8288	8300	in-frame stop codon (env)
FT	CDS	8301	8394	tat gene product (AA 100-130)
FT	CDS	8301	8551	art gene product (AA 24-108) (8301 is 2nd base in codon)
FT	SITE	8540	8757	ORF (F)

FT	CDS	9572	9357	F-gene product (AA 1-262)
FT	SITE	9587	9371	LTR (U3)
FT	RTT	9587	9349	3' LTR
FT	PHM	9368	9377	enhancer-like sequence
FT	PRM	9382	9391	enhancer-like sequence
FT	SITE	9392	9402	conserved sequence
FT	SITE	9404	9412	conserved sequence
FT	SITE	9414	9424	conserved sequence
FT	SITE	9472	9646	LTR (R)
FT	PRM	9491	9496	pot. promoter sequence

XX
SQ Sequence 9646 BP: 3290 A: 1817 C: 2387 G: 2152 T: 0 other:

Initial Score	=	280	Optimized Score	=	1247	Significance	=	0.00
Residue Identity	=	54%	Matches	=	1433	Mismatches	=	897
Gaps	=	281	Conservative Substitutions	=			=	0

```

X          10          20          30          40          50          60
AAGAG-----GAGAGACAGTGGCAATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGA
| | |      | | |      | | |      | | |      | | |      | | |      | | |
ATGGGTTGTCTTGGAATCAGCTGC--TTATCGCCATCTTG-CTTCTAAGTGTCTATGG-GAT-TTATTGTA
6090      6100      6110      6120      6130      6140      6150

          70          80          90          100         110         120         130
---AATGGGGCACCATGCTCCTTGGGATATTGATGATCTGTAG---TGCTACAGAAAAATTGTGGGTCACAG
| | |      | | |      | | |      | | |      | | |      | | |      | | |
TTCAATATGTCA-CAGTCT-TTATGGGTGTACCAGCT-TGGAGGAATGCGACAATTCCCCTCTTCTGTGCAA
6160      6170      6180      6190      6200      6210      6220

          140         150         160         170         180         190         200
TCTATTATGGGTACCTGTGTGGAGGAAGCA--ACC-ACCCTCTATTTTG-TGCATCAGATGCTAAAGCA
| | |      | | |      | | |      | | |      | | |      | | |      | | |
CCAGGAATAGGATACCT-TGGGGAACAACCTCAGTGCCTACCAGATAATGATGATTATTCAGA-ATTGGCCCT
6230      6240      6250      6260      6270      6280      6290

          210         220         230         240         250
T-ATGATACAGAGGTACATAATGTTTGG-----GCCAC---ACATGC---CTGTGTACCCACAG--ACC
| | |      | | |      | | |      | | |      | | |      | | |      | | |
TAATGTTACAGAAAGCTTTGATGCTTGGGAGAATACAGTCACAGAACAGGCAATAGAGGACGTATGGCAACT
6300      6310      6320      6330      6340      6350      6360

          260         270         280         290         300         310         320
C---CAACCCACAGGAAGTATATTG-GTAAA TGTGACAGAAATTTTAAAC-TGTG-GAAAAATGACATGG
| | |      | | |      | | |      | | |      | | |      | | |      | | |
CTTTGAGACCTCAATAA--AGCCTTGTGTAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGT
6370      6380      6390      6400      6410      6420      6430

          330         340         350         360         370         380         390
TAGAACAGATGCATGAGGATATAATCAGTTTAT-GGGATCAA-AGCCTAAAGCCATGTGTAAAATTAACCCC
| | |      | | |      | | |      | | |      | | |      | | |      | | |
GAG-ACAGATAAATGGGGAT-TGACAAAATCATCAACAACAACAGCATCAA--CAACAACAACAACAACAGC
6440      6450      6460      6470      6480      6490      6500

          400         410         420         430         440         450         460
ACTCTGTGTTAGTTTAAAGTGGCACTGATTTGGGGGAATGCTACTAATACCAATACTAG-TAATACCAATAGTA
| | |      | | |      | | |      | | |      | | |      | | |      | | |
AAAATCAG-TAGAGACAAGAG-AC--AT--AGTCAATGAGACTAGT-CCTTGTGTAGTTTCATGATAATTGCA
6510      6520      6530      6540      6550      6560      6570

          470         480         490         500         510         520         530
GTATCGGGGAA-ATGA-TGATGGAGAAAGGAGAGATHAAAAACTGCTCTTTCAATATCAGCACAGNATAAG
| | |      | | |      | | |      | | |      | | |      | | |      | | |
CAGGCTTGGAGACAGAGCCAAATGATAAGCTGTAAATTCAACATGACAGGGTTAAAAAGAG-ACAAGAAAAAG
6580      6590      6600      6610      6620      6630      6640

          540         550         560         570         580         590

```


[illegible]

[illegible]

```

TTGCTTTTAACATAACAAATTGGCTGTGGTATATA-AAAAT---ATTGATAATGATAGTAGGAGGCTTGGTAG
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
TTGCTTTTGACCTTACTTCTTGGATAAAGTATATACAATATGGAATTTAT-ATAATTGTAGGAG---TAATAC
8140      8150      8160      8170      8180      8190      8200

      2120      2130      2140      2150      2160      2170      2180
GTTTAAAGAATAGTTTTTGTCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCA---TTATCGT
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
TGTTAAGAATAGTGATCTATATAGTACAAATGCTAGCTAGGTTAAGACAGGGGTATAGGCCAGTGTTCTC-T
8210      8220      8230      8240      8250      8260      8270

      2190      2200      2210      2220      2230
TTCAGACCCACCT-----CCCAACCCCGAGGGGACCCGAC-AGGCC--C-----GAAGGAATAGAA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
TCCCAACCCCTCTTATTTCCAGTAGACCCATACCCAACAGGATCCGGCTCTGCCAACCAAGAAGGCCAAAAAA
8280      8290      8300      8310      8320      8330      8340

      2240      2250      2260      2270      2280      2290      2300
GAAGCAAGGTGGAGAGAGAGACAGAGACAGATC----CATTCGATTAGTGAACGGATCCTTAGCACTTATCTG
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
GGAGACGGTGGAGGGCAGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATATTCATTTTC-CTGATCCG
8350      8360      8370      8380      8390      8400      8410

      2310      2320      2330      2340      2350      2360      2370
GGACGATCTGGGAGC---CTTG---TGCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTA-A
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
---CCAAGTATAGGCTCTTGACTTGGCTATTTCAGCAA-----CTGCAGAACCTTGCTATCGAGAGCATA
8420      8430      8440      8450      8460      8470      8480

      2380      2390      2400      2410      2420      2430      2440
CGAGGATTGTGGTACTTCTGGGACGCGAGGGGTGGGAAGCCCT-CAATATTGGTGGGAATCTCCTACAGTAT
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CCA-GATCCTCGACCAATATTCAGAGGCTCTCTGCGACCCCTACGGAGAATTCGAGAA-GTCCT-CAGGCT
8490      8500      8510      8520      8530      8540      8550

      2450      X
TGGAGTCA-GGAAGTAAAG
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
TGAAGTACCTACCTACAA
8560      8570

```

8. KUNZ-158-FL32L SEQ

HIV2RODX Human immunodeficiency virus type 2 ROD isolate RN

```

ID  HIV2RODX  standard) RNA; 5671 BP.
XX
AC  X05291)
XX
DT  04-JUN-1987 (annotation)
XX
DE  Human immunodeficiency virus type 2 ROD isolate RNA genome
DE  (HIV-2)
XX
KW  acquired immune deficiency syndrome; art gene; env gene; f gene;
KW  gag gene; pol gene; q gene; r gene; tat gene.
XX
OS  Human immunodeficiency virus type 2
OS  ROD isolate
OC  Viridae; ss-RNA enveloped viruses; Retroviridae.
XX
RN  [1] (bases 1-5671)
RA  Alison M.
RT  ;
RL  Submitted (03-JUN-1987) on tape to the EMBL Data Library by:
RL  Marc Aliou, Unite d'Ecologie virale, and CNRS U01157, Institut

```


GGGTAGAAYTGGGAGACCATCTCTCTTGGGATATTG-----ATG-ATCTGTAGTGCTACAGAAAAATTG
 TATT--AGCTAGTGG---TTG---CTTAGTATATTGCACCCAATATGTAAGTGT-TTTCTATGGCGTACCCA
 6180 6190 6200 6210 6220 6230
 130 140 150 160 170 180
 TG-AGTCAAGTCTATTATGEGGTACCTGTGTGGAAGGAAGCAACCACCACTCTATTTTGTG---CAT-CAG
 CGTGGAAAAATGCAACCATTCCCTCTTTTGTGCAA-CCAGAAATAGGGA---TACTTGGGGAACCATACAG
 6240 6250 6260 6270 6280 6290 6300
 190 200 210 220 230 240 250
 ---ATGCTAAGGCAT-ATGATACAGAGGTACATAATGTTTG--GGCCACACATGCCTGTGTACCCACAGACC
 TGCTTGCCTGACAATGATGATTATCAGG-AAATAA-CTTTGAATGTAACAGAGGCTTTTG-ATGCATGGAAT
 6310 6320 6330 6340 6350 6360 6370
 260 270 280 290 300 310 320
 CCAAC-CCACAAGAGTAGTATTGGTAATG--TGACAGAAAAATTTTAACATGTGGAAAAATGACATG-GTA
 AATACAGTAAACAGCAAGCAATAGAGATGTCTGGCATCTATTCGAGACAT---CAATAAAACCATGTGTC
 6380 6390 6400 6410 6420 6430 6440
 330 340 350 360 370 380 390
 GAACAGATGCATGAGGATATATCAGTTTATGGGATCAAAGCCTAAGC--CATGTGTAAAATTAACCCAC
 AAACAAACACCTTTATGTGT-AGCAATGAAATGCAGCAGCAGAGAGCAGCACAGGGAA---CAACACAAC
 6450 6460 6470 6480 6490 6500 6510
 400 410 420 430 440 450 460
 TCTGTGTTTGTTTAAGGTGCACTGATTTGGGGAAATGCTACTAATACCAATA-CTAG---TAA-TACCAATAG
 -CTCAAGAGACACAA---GCACAACCACAACCACACCCAC--AGACCAGGAGCAAGAGATAAGTGAGGATAC
 6520 6530 6540 6550 6560 6570
 470 480 490 500 510 520
 TAGTAGCG---GAGAAATGA--TGATGGAGAAAGGAGAGATAAAAAACTGCTCTTTCAATATCAGCACAGN
 TCCAAGCGACGCGCAGACAACTGCT-CAGGATTGGGAGAGGAAGAAACGATC---AATTGCCAGTTCAA-T
 6580 6590 6600 6610 6620 6630 6640
 530 540 550 560 570 580 590
 ATAGAGG-TAAGCTGCAG-AGCAATATGCATTTTTTTTATAAACTTGATATAATACCAATAGATAATGATA
 ATGACAGGATTAGAAAGAGATAGGAAAAACAGTATAAT-GAAACATGGTA---CTCAAAGATGTGGTTT
 6650 6660 6670 6680 6690 6700 6710
 600 610 620 630 640 650 660
 CTACCAGCATTACGTTGACAAGTTGTAAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGC
 GTGAGACAGATATAGCACAA---ATCAGACC-CAGT-GTTACATGAACCATTTGC-AACACATC-----AGT
 6720 6730 6740 6750 6760 6770
 670 680 690 700 710 720 730 740
 CAATCCCATACATTATTGTGCCCGGECTGGT TTTGCGATCTAAAATGTAATAATA-AGACGTTCAATGGA
 CA--TCACAGG-ATCA-TGTGACAAG--CACTATTGGGATGCTATAAGGTTTAGATACTGTGCACCACCGGG
 6780 6790 6800 6810 6820 6830
 750 760 770 780 790 800
 ACAGGACC-ATGTACA---ATGTGAGGACAGTA--CA---ATGTACACA--TGGAATTAGGCCAGTAGTA
 TTATGCCCTTTTAAAGATGTAATGATACCA-ATTATTCAGGCTTTGCACCCAACCTGTTCTAAAG-TAGTAGCT
 6840 6850 6860 6870 6880 6890 6900
 810 820 830 840 850 860

TCAC-TCAC--TCTG-TTGAATGGCAGCTCAGCAGAAGAAGAGGGTAGTAATTAGATCTGCCAATTTTCAC
TCTTCATGACACAGGATGATGGAACGCAAACTTCCACATG---GTTTGGCTTTA-A--TGGCACTAGAGC
6910 6920 6930 6940 6950 6960 6970

870 880 890 900 910 920 930
AGACATGCTAAGACCA--TATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCCAACAAC
AGABAATAGAACATATCTATTGTCATGTCAGAGATAA---TAGAACT--ATCAT-CA-GCTTAAACA--
6980 6990 7000 7010 7020 7030

940 950 960 970 980 990 1000
AATACAGGA--AAATATCCGTATCCAGAGGGGACCAGGGA--GAGCATTGTGTAC-AATAGGAAAAAT--
AATATTATATCTCAGTTTGCATTGTAAAGAGG---CCAGGGAATAAGACAGTGAACAAATAATGCTTATGT
7040 7050 7060 7070 7080 7090 7100

1010 1020 1030 1040 1050 1060 1070
-ACCAATATGAGACAGGACATTTGTAACATTAGTAGAGCAAAATGCAATGCCACTTTAAACA----GATA
CAGACATGTTTCACTCCCACT---AC--CAGCCGATCAATA---AAAGACCCAGACAAGCATGGTGCTG
7110 7120 7130 7140 7150 7160

1080 1090 1100 1110 1120 1130
G--CTA--ACCAATTAAGAGA--ACAAT-TTGGAAAT-AATAAAACAATAATCTTTAAGCAATCCTCAGG---
GTTCAAGAGCAATGGAAGAGGAGGAGGTGAAGGAAACCT---TGCAAAACATCC-CAGGTAT
7170 7180 7190 7200 7210 7220 7230

1140 1150 1160 1170 1180 1190
AGGGACCCAGCAATTTGTAAGGACAGTTTTTAATTGTGGAG-----GGGAATTTTTC-TAC-----TGTA--
AGACCAAC--AAT-GACAAACCAATATTAGCTTTGCAGCGCCAGGAAAGGCTCAGACCCAGAAGTAGC
7240 7250 7260 7270 7280 7290 7300

1200 1210 1220 1230 1240 1250
ATTCA--ACAC-AAGTGTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAAT-AACAC
ATAATGTGTGACTACTGCAAGGAG--AGTTTCTCTACTGCAACAT-GACT--TGGTTCTCAATTGGATAG
7310 7320 7330 7340 7350 7360 7370

1260 1270 1280 1290 1300 1310 1320
TGAAGGAATGACA--CAATGACACTCCCATGCGAGAATAAAACAATTTATAACATGTGGCAGGAAGTAGG
AGAA-TAAGACACACCAATTTATGC-ACCGTGCCATATAAAGCAAATAATTAACACATGGCATAAGGTAGG
7380 7390 7400 7410 7420 7430 7440

1330 1340 1350 1360 1370 1380 1390
AAAAGCAATG--TATGCCCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAATATTACAG--GGCTGC-TA
--GAGAAATGATATTTTCCTCCCA---GGGA-AGGGGAGCTGTCTCTGCAACT-CAACAGTAACCAGCATA
7450 7460 7470 7480 7490 7500

1400 1410 1420 1430 1440 1450
TTACACAGAGATGG--TGGTAATAACAACAATGGGTCCGAGAT---CTTCAGACCTGGAGGAGG-AGA--TA
ATTCCTA-ACATGACTGGCA--ACAATAATCAGACAAACATTACCTTTAGTGC-AGAGGTGGCAGAACTA
7510 7520 7530 7540 7550 7560 7570

1460 1470 1480 1490 1500 1510 1520
TGAGGAGAC-ATGGGAGAGTGAATTATATAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCC
T-----ACGATTTGAGTTGGGAGAT-TATAAT-----TGGTAGAAATAACACCAATTGGCTTCGCACCT
7580 7590 7600 7610 7620 7630

1530 1540 1550 1560 1570 1580 1590

[illegible]

873 TTTTAAATTTATGAGGGAAATTTTCTACGTAAATTCACACAACTGGTTTAAATAGTACTTGGTTTA--A

1 1111 1111 11 11111111 11 11 11 11 11 11 11 11
 TGTGACAAATTTCAGAGGAGAGTTTCTCTACTGTAA---AA-TGAATTGGTTTCTA-AATTGGGTAGAGGA
 N 10 20 30 40 50 60
 1230 1240 1250 1260 1270 1280 1290
 TACTACTTGGACTACTGAAGGGTCAAATAACACTGAAGGAAGTGAACACAATCACACTCCCATGCAGAATAAA
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 TGGGATGTAPGACCCAGAGGCCAAGGA-AC----GGCATAGAGGAATTAC-GTGCCGTGTCATATTAG
 70 80 90 100 110 120 130
 1300 1310 1320 1330 1340 1350 1360
 ACAATTTATTAACATGTGGCTGGAGGTAGGAAGAGCAATG--TATGCCCTCCCATCAGCGGACAAATTAGA
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 ACAATAATCAACACTTGGCTTAAAGTAGGCAAA--AATGTTTATTTGCCTCCAAGAGAGGGAGACCTCAGG
 140 150 160 170 180 190 200
 1370 1380 1390 1400 1410 1420 1430
 TGTTCATCAATATTACAGGCTCTATTAAACAAGAGATGGTGGTAATAACAACAATGGGTCCGAGATCTTC
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 TGTACTCCACAGTGAACAGTCTCATAGCAACATAGATTGGACTGATGGAAACCA---AACTAATATCACC
 210 220 230 240 250 260 270
 1440 1450 1460 1470 1480 1490 1500
 AGACCTG---GAGGAGG-AGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATT
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 ATGAGTGCAGAGCTGGCAGAGCT---GTATCGATTGGAGTTGGGAGAT-TATAAAT-----TAGTAGAGATN
 280 290 300 310 320 330
 1510 1520 1530 1540 1550 1560 1570
 GAACCATTAGGAGTACACCCACCAAGGCAAGAG----AAGAGTGGT-GCA--GAGAGAAAAAGAGCAGT
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 NNNNNNNNNNNNNNNNNNNNNCCCCCAGAGTGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAG-AG-
 340 350 360 370 380 390 400
 1580 1590 1600 1610 1620 1630 1640
 GGGAAATAGGAGCTTTGTTCTTGGTCTT-TGGGAGCAGCAGGAAGCACTATGGGCGCACGGTCAATGACGC
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 GGGCTTTTGTGCTAGGCTTCTTGGCTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCNNNNNNNN
 410 420 430 440 450 460 470
 1650 1660 1670 1680 1690 1700 1710
 TTAGGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAAATTTGCTGAGGGCTATTGAGG
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 NNDGCGCTCAGTCCCGAGCTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGGTCAAGA
 480 490 500 510 520 530 540
 1720 1730 1740 1750 1760 1770 1780
 CGCAACAGCATCTGTTGCAATCAGAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAA
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 GACACCAAGAAATTGTTGCGACTGACCGTCTGGGGAACAAGAACCTCCAGACTAGGGTCACTGCCATCGAGA
 550 560 570 580 590 600 610 620
 1790 1800 1810 1820 1830 1840 1850 1860
 GATTCTTAAGGATCAACAGTCTCTGGGGATTGGGGTTGCTCTGGAAACTCATTTGCACCACTGCTGTGC
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 AGTACTTAAGGAGCCAGGCGAGCTGAATGCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTAC
 630 640 650 660 670 680 690
 1870 1880 1890 1900 1910 1920
 CTGAG---ATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTGGGAATAACATGACCTGGATGGAGTGGG
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 CAGGCCAGATGAGGT---CTAACCCAGACTGGAACA-AT---GA-TA-----CTTGGCAAGAGTGGG
 700 710 720 730 740 750
 1930 1940 1950 1960 1970 1980 1990 2000
 GAGGAGAGCTTGGTGAATTTGGAAGGATCGCAAAACCAGCAAGGAAAAAG

[illegible]

Kun3
07/158452
Claim 32
Genbank

Results file kunz-158-cl32nih.res made by sheppard on Thu 8 Mar 90 11:09:14-PST.

Query sequence being compared: KUNZ-158-CL32. SEQ
Number of sequences searched: 31228
Number of scores above cutoff: 38

Results of the initial comparison of KUNZ-158-CL32. SEQ with:
Data Bank : GenBank 62.0, all entries

100000-
--
N --
U50000-
M --
B --
E --
R *
--*
O --
F10000-
--
S --
E 5000-
Q --
U --
E --
N --
C --
E --
S 1000-
-- *
--
500-
--
--
--
--
--

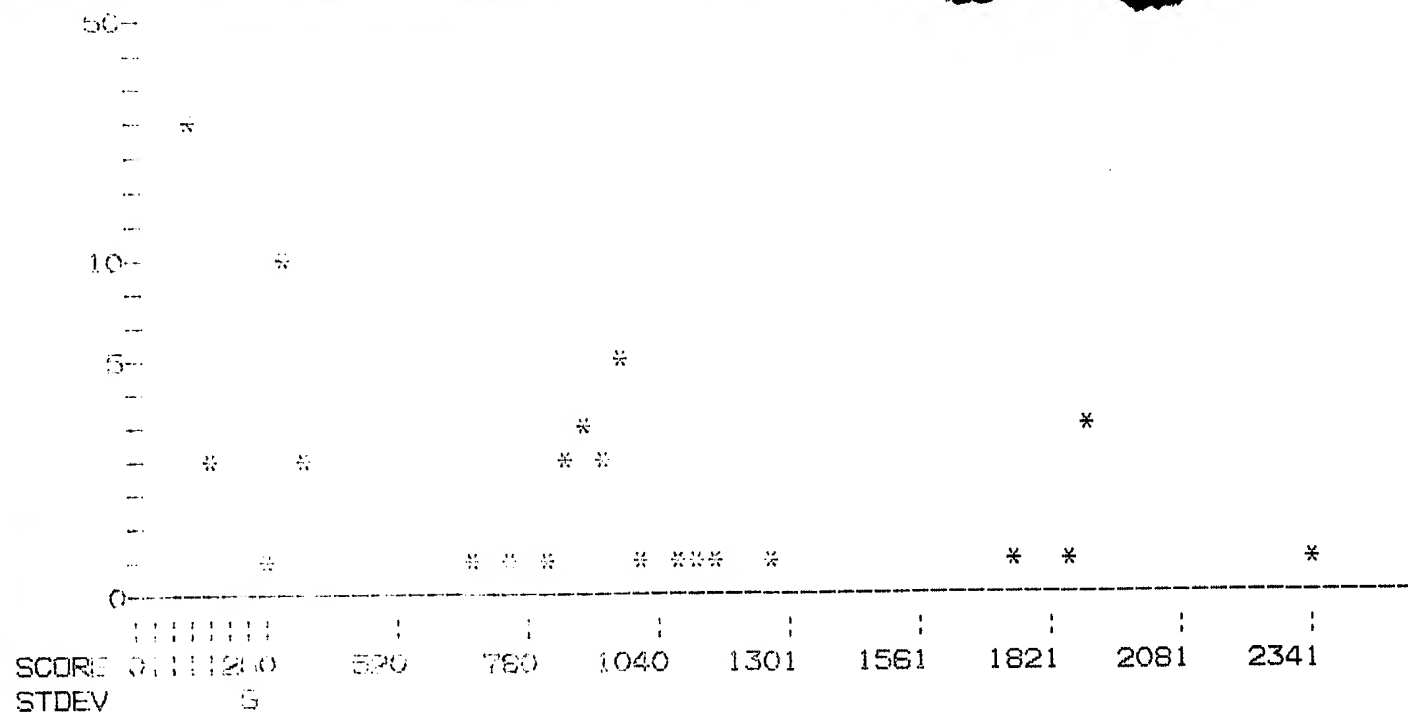
2110 2120 2130 2140 2150 2200 2210 2220 2230
TATGACAGGATATTCACCATTTATCTTTTACAGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGG
|||||
TATGACAGGATATTCACCATTTATCTTTTACAGACCCACCTCCCAATCCCGAGGGGACCCGACAGGCCCGAAGG
X 10 20 30 40 50 60 70

2240 2250 2260 2270 2280 2290 2300
AATAGAAAGATGAGGATGGAGAGAGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTAT
|||||
AATAGAAAGATGAGGATGGAGAGAGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTAT
60 80 100 110 120 130 140

2310 2320 2330 2340 2350 2360 2370
CTGGGACGATCTGCGGAGCCTGATGCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGA
|||||
CTGGGACGATCTGCGGAGCCTGATGCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGA
150 160 170 180 190 200 210

2380 2390 2400 2410 2420 2430 2440
GGATTGTGGAAGTTCTGGGAGCGCAGGGGGTGGGAAGCCCTCAAATATTGGTGGGAATCTCCTACAGTATTGGA
|||||
GGATTGTGGAAGTTCTGGGAGCGCAGGGGGTGGGAAGCCCTCAAATATTGGTGGGAATCTCCTACAGTATTGGA
250 260 270 280

2450 X
GTGAGGAACTAGAG
|||||
GTGAGGAGCTAAGG
290 300



PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	147		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	37	37	19.85

Times:	CPU	Total Elapsed
	00:52:54.08	02:09:12.00

Number of matches:	37183850
Number of sequences searched:	31228
Number of scores above cutoff:	38

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Init.	Opt.	Length	Score	Score	Sig.	Frame
---------------	-------------	-------	------	--------	-------	-------	------	-------

**** 116 standard deviations above mean ****

1	144954138	Human immunodeficiency virus t	9229	2341	2456	116.09	0
---	-----------	--------------------------------	------	------	------	--------	---

**** 22 standard deviations above mean ****						
2. HIVPV22	Human immunodeficiency virus t	9770	1877	2190	92.71	0
3. HIVHXB3	Human immunodeficiency virus t	3156	1873	2186	92.51	0
4. HIVBH102	Human immunodeficiency virus t	8932	1872	2186	92.46	0
**** 51 standard deviations above mean ****						
5. HIVHXB205	Human immunodeficiency virus t	9718	1858	2174	91.75	0
**** 55 standard deviations above mean ****						
6. HIVNL43	Human immunodeficiency virus t	9709	1729	2181	85.25	0
**** 60 standard deviations above mean ****						
7. HIVEL106	Human immunodeficiency virus t	9176	1246	1904	60.91	0
**** 56 standard deviations above mean ****						
8. HIVSC	Human immunodeficiency virus t	4273	1159	2151	56.53	0
**** 54 standard deviations above mean ****						
9. HIVH3BH3	Human immunodeficiency virus t	3563	1112	1773	54.16	0
**** 51 standard deviations above mean ****						
10. HIVZ321	Human immunodeficiency virus t	3457	1066	1970	51.85	0
**** 48 standard deviations above mean ****						
11. HIVMND1	Human immunodeficiency virus t	9738	996	2223	48.32	0
**** 47 standard deviations above mean ****						
12. HIVZ225	Human immunodeficiency virus t	9081	983	1920	47.66	0
13. HIVZ6	Human immunodeficiency virus t	5159	983	1924	47.66	0
14. HIVTH34	Human immunodeficiency virus t	2903	975	1385	47.26	0
15. HIVJY1	Human immunodeficiency virus t	2653	972	1936	47.11	0
**** 45 standard deviations above mean ****						
16. HIVCDC92	Human immunodeficiency virus t	3373	932	2209	45.09	0
**** 44 standard deviations above mean ****						
17. HIVMAL	Human immunodeficiency virus t	9229	916	2060	44.29	0
**** 42 standard deviations above mean ****						
18. HIVRFEHV	Human immunodeficiency virus ty	2622	887	1587	42.83	0
19. HIVRF	Human immunodeficiency virus t	9128	887	1587	42.83	0
20. HIVSF203	Human immunodeficiency virus t	9737	883	1960	42.63	0

The scores below are sorted by optimized score.

Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. HIVSRU15	Human immunodeficiency virus t	9229	2341	2456	0.00	0
2. HIVMND1	Human immunodeficiency virus t	9738	996	2223	0.00	0
3. HIVCDC92	Human immunodeficiency virus t	3373	932	2209	0.00	0
4. HIVPV22	Human immunodeficiency virus t	9770	1877	2190	0.00	0
5. HIVBH102	Human immunodeficiency virus t	8932	1872	2186	0.00	0
6. HIVHXB3	Human immunodeficiency virus t	3156	1873	2186	0.00	0
7. HIVNL43	Human immunodeficiency virus t	9709	1729	2181	0.00	0
8. HIVHXB205	Human immunodeficiency virus t	9718	1858	2174	0.00	0
9. HIVSC	Human immunodeficiency virus t	4273	1159	2151	0.00	0
10. HIVMAL	Human immunodeficiency virus t	9229	916	2060	0.00	0
11. HIVZ321	Human immunodeficiency virus t	3457	1066	1970	0.00	0
12. HIVSF203	Human immunodeficiency virus t	9737	883	1960	0.00	0
13. HIVJY1	Human immunodeficiency virus t	2653	972	1936	0.00	0
14. HIVZ6	Human immunodeficiency virus t	5159	983	1924	0.00	0
15. HIVZ225	Human immunodeficiency virus t	9081	983	1920	0.00	0
16. HIVEL106	Human immunodeficiency virus t	9176	1246	1904	0.00	0
17. HIVH3BH3	Human immunodeficiency virus t	3563	1112	1773	0.00	0
18. HIVRFEHV	Human immunodeficiency virus ty	2622	887	1587	0.00	0
19. HIVRF	Human immunodeficiency virus t	9128	887	1587	0.00	0
20. HIVSRU15	Human immunodeficiency virus t	3600	869	1479	0.00	0

HIVBRU001

Human immunodeficiency virus type 1, isolate BRU,

LOCUS HIVBRU001 9229 bp ss-RNA VRL 15-JUN-1989
 DEFINITION Human immunodeficiency virus type 1, isolate BRU, complete genome (LAV-1).
 ACCSSION K02015
 KEYWORDS TAR protein; TAT protein; acquired immune deficiency syndrome; complete genome; env gene; gag gene; long terminal repeat; pol gene; polyprotein; provirus; reverse transcriptase; trans-activation.
 SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate BRU (LAV-1), proviral DNA clone lambda-J18.
 ORGANISM Human immunodeficiency virus type 1
 Viruses; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.
 REFERENCE 1 (bases 1 to 9229)
 AUTHORS Nathanson, C.L., Sonigo, P., Dams, O., Cole, S. and Alizon, M.
 TITLE Nucleotide sequence of the AIDS Virus, LAV
 JOURNAL Cell 40, 5-17 (1985)
 STANDARD full staff_review
 REFERENCE 2 (bases 1712 to 1749; revision of [1])
 AUTHORS Alizon, M., Vain-Hobson, S., Montagnier, L. and Sonigo, P.
 TITLE Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isolates from African patients
 JOURNAL Cell 42, 63-74 (1986)
 STANDARD full staff_review
 COMMENT The original LAV, sometimes called LAV-1 to distinguish it from HIV-2 (LAV-2), is now referred to as HIV-1bru. An infectious clone of this virus has been constructed by Keith Peden, Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205 (301) 955-3652. HIVNL43 is also an infectious clone having for its 3' half a clone of the BRU isolate. Acquired immune deficiency syndrome (AIDS) is caused by a retrovirus known by several different names, probably representing two separate strains: human T-cell lymphotropic virus-III (HTLV-III) and lymphadenopathy-associated virus (LAV) are thought to be one strain, and AIDS-associated retrovirus type 2 (ARV-2) the other. All three viruses, whose sequences do not differ by more than about 8%, are believed to belong to the retroviral subfamily Lentiviridae, or "slow" viruses. For the details of the annotation and for other pertinent references, see the HIV reference entry.

FEATURES	from	to/span	description
pept	336	1874	gag polyprotein
pept	1631	4678	pol polyprotein (NH2-terminus uncertain; AA at 1631)
pept	4623	5201	vif protein
pept	5141	5431	vpr protein
pept	5412	5626	tat protein, exon 2 (first expressed exon)
	7572	8017	tat protein, exon 3 (AA at 7973)
pept	5551	5626	rev protein, exon 2 (first expressed exon)
	7572	8246	rev protein, exon 3 (AA at 7974)
pept	5643	5989	vpu protein
pept	5803	6383	envelope polyprotein
pept	6350	6010	nef protein
pre-msg	1	5229	genomic mRNA
pre-msg	1	9229	tat, rev, nef subgenomic mRNA
IVS	250	5358	tat, rev, nef subgenomic mRNA intron 1
IVS	5627	7571	tat cds intron 2
IVS	5627	7571	rev cds intron 2
IVS	5627	7571	tat, rev, nef subgenomic mRNA intron 2
LTR	(1	130	5' LTR
LTR	9678) 9229	3' LTR
rpt	(1	97	R repeat 5' copy
rpt	9133	9229	R repeat 3' copy
binding	152	193	primer (lys-tRNA) binding site

revision 1712 1749 attttcttcagagcagaccagagccaacagccccaccag in [2];
ag in [1]
signal 3205 3210 mRNA polyadenylation signal
BASE COUNT 3268 a 1656 c 2232 g 2052 t
ORIGIN Cap site of genomic RNA.

Initial Score = 2341 Optimized Score = 2456 Significance = 0.00
Residue Identity = 99% Matches = 2456 Mismatches = 3
Gaps = 1 Conservative Substitutions = 0

```
X      10      20      30      40      50      60      70
AAGAGCAGAAGACAGTGGCAATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAAATGGG
|||||
AAGAGCAGAAAGACAGTGGCAATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAAATGGG
X  5790      5800      5810      5820      5830      5840      5850

      80      90      100      110      120      130      140
GCACCATGCTCCTTGGGATATTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGG
|||||
GCACCATGCTCCTTGGGATATTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTACAGTCTATTATGGGG
5860      5870      5880      5890      5900      5910      5920

      150      160      170      180      190      200      210
TACCTGTGTGGAAGGAAGCAACCACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTAC
|||||
TACCTGTGTGGAAGGAAGCAACCACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTAC
5930      5940      5950      5960      5970      5980      5990

      220      230      240      250      260      270      280
ATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGA
|||||
ATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGA
6000      6010      6020      6030      6040      6050      6060      6070

      290      300      310      320      330      340      350      360
CAGAAAAATTTTAACATGTGGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATC
|||||
CAGAAAAATTTTAACATGTGGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATC
6080      6090      6100      6110      6120      6130      6140

      370      380      390      400      410      420      430
AAAGCCTAAAGCCATGTGTAAATTAACCCCACTCTGTGTTAGTTTAAAGTGCCTGATTTGGGGAATGCTA
|||||
AAAGCCTAAAGCCATGTGTAAATTAACCCCACTCTGTGTTAGTTTAAAGTGCCTGATTTGGGGAATGCTA
6150      6160      6170      6180      6190      6200      6210

      440      450      460      470      480      490      500
CTAATACCAATAGTAATAACCAATAGTAGTAGCGGGGAATGATGATGGAGAAAGGAGAGATAAAAACT
|||||
CTAATACCAATAGTAATAACCAATAGTAGTAGCGGGGAATGATGATGGAGAAAGGAGAGATAAAAACT
6220      6230      6240      6250      6260      6270      6280

      510      520      530      540      550      560      570
GCTCTTTCAATATCAGCACAAGGATAAGAGGTAAAGGTGCAGAAAGAATATGCATTTTTTTATAAACTTGATA
|||||
GCTCTTTCAATATCAGCACAAGGATAAGAGGTAAAGGTGCAGAAAGAATATGCATTTTTTTATAAACTTGATA
6290      6300      6310      6320      6330      6340      6350

      580      590      600      610      620      630      640
TAATACCAATAGTAATAACTAACAAGCTATACCTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCT
|||||
TAATACCAATAGTAATAACTAACAAGCTATACCTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCT
6360      6370      6380      6390      6400      6410      6420      6430

      650      660      670      680      690      700      710      720
GTCCCAAGTAAATTTTGGGCTGCTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAATGTA
```

6440 6450 6460 6470 6480 6490 6500
 GTCTTAAAGGATCTCTTGAAGCCAAATCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAATGTA
 6510 6520 6530 6540 6550 6560 6570
 ATAATAAGACGTTCAATCGAAGACAGAACATGTTCAAATGTTCAGCACAGTACAATGTACACATGGAATTAGGC
 6580 6590 6600 6610 6620 6630 6640
 CASTAGTATCAACTCAACTGCTGTGAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGCCAATT
 6650 6660 6670 6680 6690 6700 6710
 TCACAGACAAATGCTAAACCCATTAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACA
 6720 6730 6740 6750 6760 6770 6780 6790
 ACAATACAGAGAAAGATATCCGTATCCAGAGGGGACCCAGGGAGAGCATTGTGTTACAATAGGAAAAATAGGAA
 6800 6810 6820 6830 6840 6850 6860
 ATATGAGAAAGAGACCTTCTAACAATAGTACAGCAGAAATGGAATGCCACTTTAAAACAGATAGCTAGCAAAT
 6870 6880 6890 6900 6910 6920 6930
 TAATGAGAAAGAGACCTTCTAACAATAGTACAGCAGAAATGGAATGCCACTTTAAAACAGATAGCTAGCAAAT
 6940 6950 6960 6970 6980 6990 7000
 CGCACAGTCTTTAATTCAGGAGGAGAAATTTCTACTGTAATTCACACAACTGTTTAATAGTACTTGGTTTA
 7010 7020 7030 7040 7050 7060 7070
 ATACTACTTGGTAAATGAGGAGGAGAAATTCACACAACTGTTTAATAGTACTTGGTTTA
 7080 7090 7100 7110 7120 7130 7140 7150
 AAGTATTATATAACATGAGGAGGAGAAATTCACACAACTGTTTAATAGTACTTGGTTTA
 7160 7170 7180 7190 7200 7210 7220
 GTTCAATCAATATAGTACAGGAGGAGAAATTCACACAACTGTTTAATAGTACTTGGTTTA
 7230 7240 7250 7260 7270 7280 7290 7300
 GTTCAATCAATATAGTACAGGAGGAGAAATTCACACAACTGTTTAATAGTACTTGGTTTA

[illegible]

[illegible]

IVS	6060	9385	tat, rev, nef subgenomic mRNA intron 2
LTR	1	633	5' LTR
LTR	9106	9738	3' LTR
npf	453	550	R repeat 5' copy
npf	8558	8585	R repeat 3' copy
binding	378	380	Spl binding site III
binding	387	396	Spl binding site II
binding	398	407	Spl binding site I
binding	635	652	primer (Lys-tRNA) binding site
site	3783	3785	pol cds in-frame stop codon
signal	8631	8636	mRNA polyadenylation signal
BASE COUNT	3463 a	1759 c	2344 g 2162 t
ORIGIN	Left end of viral genome		

Initial Score	=	598	Optimized Score	=	2223	Significance	=	0.00
Residue Identity	=	90%	Matches	=	2253	Mismatches	=	170
Gaps	=	58	Conservative Substitutions	=			=	0

```

X      10      20      30      40      50      60      70
AASPGCAAGACACASTGGCAATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAAATGGG
|||||
AASPGCAAGACACASTGGCAATGAGAGTGAAGG--GGATCAGGAGGAATTAT-CAG-CACTGGTGGGGATGGG
6220      6230      6240      6250      6260      6270      6280

      80      90      100     110     120     130     140
GCACCATCTCTTTCATATTTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTCTATTATGGGG
|||||
GCACCATCTCTTTCATATTTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTCTATTATGGGG
6290      6300      6310      6320      6330      6340      6350

      150     160     170     180     190     200     210
TACCTGTGTGGAAAGGAGCAACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTAC
|||||
TACCTGTGTGGAAAGGAGCAACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTAC
6360      6370      6380      6390      6400      6410      6420      6430

      220     230     240     250     260     270     280
ATAATGTTTGGGCCACACATGCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGA
|||||
ATAATGTTTGGGCCACACAGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGAATTGGTAAATGTGA
6440      6450      6460      6470      6480      6490      6500

290      300      310      320      330      340      350      360
CAGAAATTTTACATGTGGAAATGACATGCTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATC
|||||
CAGAAATTTTACATGTGGAAATGACATGCTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATC
6510      6520      6530      6540      6550      6560      6570

      370      380      390      400      410      420      430
AAAGCCTAAAGCCATGTTTAAAAATTACCCCACTCTGTGTTAGTTTAAAGTGCCTGATTTGGGGAATGCTA
|||||
AAAGCCTAAAGCCATGTTTAAAAATTACCCCACTCTGTGTTAGTTTAAAGTGCCTGATTTGAGGAATACTA
6580      6590      6600      6610      6620      6630      6640

      440      450      460      470      480      490
CTATACCAATCTAG-----TAATACCAATAGTAGTAGCGGGGAATGATGATGGAGAAAGGAGAGATAA
|||||
CTATACCAATCTAGTAGTAGTAGTAGCGGGGAATGATGATGGAGAAAGGAGAGATAA
6650      6660      6670      6680      6690      6700      6710

500      510      520      530      540      550      560      570
AAAAGTGCCTTTTCAAATATCAGCAAAAGNATAAGAGGTAAAGTGCAGAAAGAATATGCATTTTTTTATAAAC
|||||
AAAAGTGCCTTTTCAAATATCAGCAAAAGNATAAGAGGTAAAGTGCAGAAAGAATATGCATTTCTTTATAAAC
6720      6730      6740      6750      6760      6770      6780

```

570 580 590 600 610 620 630 640
TTGATATAATACGAATGATAATGATACTACCAGCTATACGTTGACAAGTTGTAAACACCTCAGTCATTACAC
TT
TTGATATAATACGAATGATAATGATACTACCAGCTATAGGTTGATAAGTTGTAAACCTCAGTCATTACAC
6790 6800 6810 6820 6830 6840 6850 6860

650 660 670 680 690 700 710
AGGCTGTGCAAAAGGTATTCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGGCTGGTTTTGCGATTCTAA
TT
AAGCTGTGCAAAAGGTATTCCTTTGAGCCAATTCCCATACACTATTGTGCCCCGGGCTGGTTTTGCGATTCTAA
6870 6880 6890 6900 6910 6920 6930

720 730 740 750 760 770 780
AATGTAAATGATAAGAGCTTCATGTCGACAGGAGCATGTACAAATGTGACGACAGTACAATGTACACATGGAA
TT
AATGTAAATGATAAGAGCTTCATGTCGACAGGAGCATGTACAAATGTGACGACAGTACAATGTACACATGGAA
6940 6950 6960 6970 6980 6990 7000

790 800 810 820 830 840 850
TTAGGCCAATGACTATCAACTCAACTGCTGTGTAATGTCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTG
TT
TTAGGCCAATGACTATCAACTCAACTGCTGTGTAATGTCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTG
7010 7020 7030 7040 7050 7060 7070

860 870 880 890 900 910 920 930
CCAAATTTCCACAGACAAATGCTAAAACCATAAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGAC
TT
AGAAATTTCACTGATTAATGCTAAAACCATCATAGTACATCTGAATGAATCTGTACAAATTAATTGTACAAGAC
7080 7090 7100 7110 7120 7130 7140

940 950 960 970 980 990 1000
CCACCAAGCAATCAAGGAAAAAGTATCCGATATCCAGAGGGGACCAGGGAGAGCATTGTGTTACAATAGGAAA--
TT
CCAACTACAAATCAAGGAAAAAGTATCATAT--AG----GACCAGGGAGAGCATTGTTATACAACAAAAAATA
7150 7160 7170 7180 7190 7200 7210

1010 1020 1030 1040 1050 1060 1070
-AATAGGAAATATGAAACAGACACATTGTAAACATTAGTAGAGCAAATGCAATGCCACTTTAAACAGATAG
TT
TAATAGGAAATATGAAACAGACACATTGTAAACATTAGTAGAGCAAATGGAATGACACTTTAAGACAGATAG
7220 7230 7240 7250 7260 7270 7280

1080 1090 1100 1110 1120 1130 1140
CTAGCAAAATTAAGAGAACAAATTTGAAATAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAG
TT
TTAGCAAAATTAAGAGAACAAATTT--AAGAATAAAACAATAGTCTTTAATCAATCCTCAGGAGGGGACCCAG
7290 7300 7310 7320 7330 7340 7350

1150 1160 1170 1180 1190 1200 1210
AAATTTGTACGACACAGTTTAAATTTGTGAGGGGAATTTTCTACTGTAATTCAACACAACCTGTTTAATAGTA
TT
AAATTTGTACGACACAGTTTAAATTTGTGAGGGGAATTTTCTACTGTAATACATCACCCTGTTTAATAGTA
7360 7370 7380 7390 7400 7410 7420

1220 1230 1240 1250 1260 1270 1280
CTTGG---TTTCTAGTACTTGG---AGTACTGAAGGGTCAAATAACACTGAAGGAAGTGACACAATCACAC
TT
CTTGGAAATGTAATAATACTTGAATAATACTACAGGGTCAAATAACAAT-----ATCACAC
7430 7440 7450 7460 7470 7480

1290 1300 1310 1320 1330 1340 1350
TCCCATGCAAGATAAAGCAATTTATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATCA
TT
TTCTATGCAAAATTAAGCAATTTATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATG
7490 7500 7510 7520 7530 7540 7550

1320 1370 1380 1390 1400 1410 1420
GCCGACAAATTGATGATTCATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAACA---AC-
|||||
AAGTAUCAAATTGATGATTCATCAAATATTACAGGGCTACTATTAACAAGAGATGGTGGTAAGGACACGGACA
7550 7570 7580 7590 7600 7610 7620

1430 1440 1450 1460 1470 1480 1490
---AATGGGTCGAGATTTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAAT
|||||
CGAACGACACCGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAAT
7630 7640 7650 7660 7670 7680 7690 7700

1500 1510 1520 1530 1540 1550 1560
ATAAAGTASTAAGAAATTGAACCATTTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAA
|||||
ATAAAGTASTAAGAAATTGAACCATTTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAA
7710 7720 7730 7740 7750 7760 7770

1570 1580 1590 1600 1610 1620 1630
AAGAGAGCACTGCGAATAGGAGCTTTGTTTCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCACGGT
|||||
AAGAGAGCACTGCGAATAGGAGCTTTGTTTCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCACGGT
7780 7790 7800 7810 7820 7830 7840

1640 1650 1660 1670 1680 1690 1700
CAATGACGCTGACGCTACAGGCGCAACAAATTATTGTCTGGTATAGTGACAGCAGCAGAACAAATTTGCTGAGGG
|||||
CAATGACGCTGACGCTACAGGCGCAACAAATTATTGTCTGGTATAGTGACAGCAGCAGAACAAATTTGCTGAGGG
7850 7860 7870 7880 7890 7900 7910

1710 1720 1730 1740 1750 1760 1770
CTATTGACGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGG
|||||
CTATTGACGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGG
7920 7930 7940 7950 7960 7970 7980

1780 1790 1800 1810 1820 1830 1840 1850
CTGTGGAAGATACCTTAAGGATCAACAGCTCCTGGGGATTGCTCTGGGAACTCATTTGCACCA
|||||
CTGTGGAAGATACCTTAAGGATCAACAGCTCCTGGGGATTGCTCTGGGAACTCATTTGCACCA
7990 8000 8010 8020 8030 8040 8050

1860 1870 1880 1890 1900 1910 1920
CTGCTGTGCTTGGAAATGCTAGTTGGAAATAATAATCTCTGGAACAGATTTGGAATAACATGACCTGGATGG
|||||
CTGCTGTGCTTGGAAATGCTAGTTGGAAATAATAATCTCTGGAATGATATTTGGAATAACATGACCTGGATGG
8060 8070 8080 8090 8100 8110 8120

1930 1940 1950 1960 1970 1980 1990
AGTGGGACGAGAAATTAACAATTACACAAGCTTAATACATTCCTTAATTGAAGAATCGCAAACCAGCAAG
|||||
AGTGGGACGAGAAATTAACAATTACACAAGCTTAATATACTCATTACTAGAAAAATCGCAAACCAACAAG
8130 8140 8150 8160 8170 8180 8190 8200

2000 2010 2020 2030 2040 2050 2060
AAAGGAATTAACCAAGAAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAAACATAACAAATT
|||||
AAAGGAATTAACCAAGAAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAAACATAACAAATT
8210 8220 8230 8240 8250 8260 8270

2070 2080 2090 2100 2110 2120 2130
GGCTCTGCTATTAACCAATATTCATAATGATACTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGTCTGTAC
|||||
GGCTCTGCTATTAACCAATATTCATAATGATACTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGTCTGTAC
8280 8290 8300 8310 8320 8330 8340

```

2140      2150      2160      2170      2180      2190      2200      2210
TTTCTATATGATATAGATTAGGTAAGGATATTCACCATTCGTTTCAGACCCACCTCCCAACCCCGAGGG
|||||
TTTCTATATGATATAGATTAGGTAAGGATATTCACCATTCGTTTCAGACCCGCCCCCAGTTCCGAGGG
8350      8360      8370      8380      8390      8400      8410

      2220      2230      2240      2250      2260      2270      2280
GACCCGACAGAGTCCGAGAGGATATAGAGAGAGAGGTGGAGAGAGAGACAGAGACAGATCCATTTCGATTAGTGA
|||||
GACCCGACAGAGTCCGAGAGGATATAGAGAGAGAGGTGGAGAGAGAGACAGAGACACATCCGGTCGATTAGTGC
8420      8430      8440      8450      8460      8470      8480

      2290      2300      2310      2320      2330      2340      2350
ACGGATCCCTAGCACTTATCTGGGACGATCTGCGGAGCCTTGTGCCTCTTCAGCTACCACCGCTTGAGAGAC
|||||
ATGGAATCTTACCAATTATCTGGGTCGACCTGCGGAGCC-TGTTCTCTTCAGCTACCACCAC---AGAGAC
8490      8500      8510      8520      8530      8540      8550

      2360      2370      2380      2390      2400      2410      2420
TTACTCTTTGATCTGTAACGAGGATTTGTGAAGTCTCTGGGACGCGAGGGGGGTGGGAAGCCCTCAAATATTGGTGG
|||||
TTACTCTTTGATCTGTAACGAGGATTTGTGAAGTCTCTGGGACGCGAGGGGGGTGGGAAGTCTCTCAAATATTGGTGG
8560      8570      8580      8590      8600      8610      8620

      2430      2440      2450      2460
AATCTCCTACAGTATTGGAGTCAGGAAGTAAAG
|||||
AATCTCCTACAGTATTGGAGTCAGGAAGTAAAG
8630      8640      8650      8660 X

```

```

pre-mag (      1      ) 3373      tat, rev, nef subgenomic mRNA
IVS      (      1      ) 34       tat, rev, nef subgenomic mRNA intron 1
IVS      303      2666      tat cds intron 2
IVS      303      2666      rev cds intron 2
IVS      303      2666      tat, rev, nef subgenomic mRNA intron 2
BASE COUNT 1174 a 526 c 809 g 805 t
ORIGIN      1 bp upstream of EcoRI site; about 3.6 kb after segment 1.

```

```

Initial Score      =      932      Optimized Score      =      2208      Significance      =      0.00
Residue Identity   =      88%      Matches              =      2241      Mismatches      =      196
Gaps              =      50      Conservative Substitutions      =      0

```

```

X      10      20      30      40      50      60      70
AASAGCAGAGAGACAGTGGCAATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAATGGG
||||| ||||||| ||||||| ||||||| ||| | | | ||||| ||| ||||| |||||
AAGAGAAAGAGACAAATGGCAATGAGAGCGAAGG-GGATCAGGAAGAATTGT-CAG-CACTTGTGGAGATGGG
X      470      480      490      500      510      520      530

      80      90      100      110      120      130      140
GCACCATGCTCTCTTGGATATTTGATGATCTGTAGTGTCTACAGAAAATTGTGGGTACAGTCTATTATGGGG
||||| ||||||| ||||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
GCACCATGCTCTCTTGGATATTTGATGATCTGTAGTGTCTGAGCAAACTTGTGGGTACAGTCTATTATGGGG
      540      550      560      570      580      590      600

      150      160      170      180      190      200      210
TACCTGTGAGGAGAGGAGACACCACTCTAATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTAC
||||| ||||||| ||||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
TACCTGTGAGGAGAGGAGACACCACTCTAATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGCAC
      610      620      630      640      650      660      670

      220      230      240      250      260      270      280
ATAATGTTTGGGCCACACATGCTTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGA
||||| ||||||| ||||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
ATAATGTTTGGGCCACACATGCTTGTGTACCCACAAACCCCAACCCACAAGAAGTAGTATTGGAAAATGTGA
      680      690      700      710      720      730      740

230      300      310      320      330      340      350      360
CAGAGAAATTTACATGTGGAAAAATGACATGATAGAACAGATGCATGAGGATATAATCAGTTTATGGGATC
||||| ||||||| ||||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CAGAGAAATTTTACATGTGGAAAAATGACATGATAGAACAGATGCATGAGGATATAATCAGCTTATGGGATC
      750      760      770      780      790      800      810

      370      380      390      400      410      420      430
AAAGCCTAAGAGCATGTGTAAAAATTAACCCCACTCTGTGTTAGTTTAAAGTGCCTGATTTGGGGAATGCTA
||||| ||||||| ||||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AAAGCCTAAGAGCATGTGTAAACTAACCCCACTCTGTGTTACTTTAAATTGCCTGATTTGAATACTAATA
820      830      840      850      860      870      880      890

      440      450      460      470      480      490      500
CTAATACCAATACTA--GTAATACCAATAGTAGTAGCGGGGAAATGATGATGGAGAAAGGAGAGATAAAAAA
||| ||| ||||||| ||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
ATACTACTAATACTACTGAAGTATCAATAAATAGTAGTTTGGGAACAACG--GGGTAAAGGAGAAATGAGAAA
      900      910      920      930      940      950      960

      510      520      530      540      550      560      570
CTGCTCTTTCAATATCAGCACAAAGNATAAGAGGTAAAGGTGCAGAAAGAATATGCATTTTTTTATAAACTTGA
||| ||||||| ||||||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CTGCTCTTTCAATATCAGCACAAAGCATAAGAGATAAGGTGCAGAGAGAATATGCATTGTTTTATAAACTTGA
      970      980      990      1000      1010      1020      1030

      580      590      600      610      620      630
TATAATACCAATAG-----ATAATGATACTACCAGC-----TATACGTTGACAAGTTGTAACACCTC
||| ||| ||||||| ||| ||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
TGTAAGAACCAATAGATGATAATAAAATACTACCAACAACACCAAAATATAGGTTGATAAATTGTAACACCTC
      1040      1050      1060      1070      1080      1090      1100

```

640 650 660 670 680 690 700
 AGTCATTACACAGGCGCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTT
 1110 1120 1130 1140 1150 1160 1170
 AGTCATTACACAGGCGCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTACCCCGACTGGTTT
 710 720 730 740 750 760 770
 TGCACTTCGAAAATGTAAATAAGACGTTCAATGGACAGGACCATGTACAAATGTCAGCACAGTACAATG
 1180 1190 1200 1210 1220 1230 1240
 TGCACTTCGAAAATGTAAACATAAGACGTTCAATGGACAGGACCATGTACAAATGTCAGCACAGTACAATG
 780 790 800 810 820 830 840
 TACACATGGAAATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCAGTCTAGCAGAAGAAGAGGTAGT
 1250 1260 1270 1280 1290 1300 1310 1320
 TACACATGGAAATTAGGCCAGTAGTGTCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGGTAGT
 850 860 870 880 890 900 910
 AATTAGATCTGCAATTTACAGGACAATGCTAAAACCATAATAGTACAGCTGAACCAATCTGTAGAAATTAA
 1330 1340 1350 1360 1370 1380 1390
 AATTAGATCTGCAATTTACAGGACAATGCTAAAACCATAATAGTACAGCTGAATGTATCTGTAGAAATTAA
 920 930 940 950 960 970 980 990
 TTGTACAAGACCCCAACCAATACAAAGAAAAAGTATCCGTATCCAGAGGGGACCAGGGAGAGCATTGTGTTAC
 1400 1410 1420 1430 1440 1450
 TTGTACAAGACCCCAACCAATACAAAGAAAAAG----GGTAAC--GCTAGGACCAGGGAGAGTATGGTATAC
 1000 1010 1020 1030 1040 1050 1060
 AATAGGAAAA--TAGGAAATATGASACAAGCACATTGTAACATTAGTAGAGCAAAATGCAATGCCACTTT
 1460 1470 1480 1490 1500 1510 1520 1530
 AACAGGAGAAATACTAGGAAATATAAGGCAAGCACATTGTAACATTAGTAGAGCACAAATGGAATAACACTTT
 1070 1080 1090 1100 1110 1120 1130
 AAAACAGATAGCTAGCAATTAAGAGAACAAATTTGGAAATAATAAAACAATAATCTTTAAGCAATCCTCAGG
 1540 1550 1560 1570 1580 1590 1600
 ACAACAGATAGCTAGCAATTAAGAGAACAAATTTGG--GAATAAAACAATAGCCTTTAATCAATCCTCAGG
 1140 1150 1160 1170 1180 1190 1200
 AGGGACCCAGAAATTGTAACGACAGTTTTTAATTGTGGAGGGGAATTTTCTACTGTAATTCACACAACT
 1610 1620 1630 1640 1650 1660 1670
 AGGGACCCAGAAATTGTAATGCACAGTTTTTAATTGTGGAGGGGAATTTTCTACTGTAATTCACACAGCT
 1210 1220 1230 1240 1250 1260
 GTTTAATAGTACTTG----GTT-----TAATAGTACTTGGAG---TACT-GAAGGGTCAAATA-ACACTGAA
 1680 1690 1700 1710 1720 1730 1740
 GTTTAATAGTACTTGGAATGTTACTAGTAATGGTACTTGGAGTGTTACTAGAAAG--CAAAAAGACACTG--
 1270 1280 1290 1300 1310 1320 1330
 GGAATGTGACATCAATCAACTCCCATGCAGAAATAAACAAATTTATAAACATGTGGCAGGAAGTAGGAAAAGCA
 1750 1760 1770 1780 1790 1800
 ----GAGACATTATCAGACTCCCATGCAGAAATAAACAAATTTATAAACAGGTGGCAGGTTGTAGGAAAAGCA
 1340 1350 1360 1370 1380 1390 1400
 ATGTTTGGCCCTCCCATCAAGCGGACAAATTAGATGTTTCATCAAATATTACAGGGCTGCTATTAACAAGAGAT
 1810 1820 1830 1840 1850 1860 1870 1880
 ATGTTTGGCCCTCCCATCAAGGACTAATTAGATGTTTCATCAAATATTACAGGGCTGCTATTAACAAGAGAT

[illegible]

```

2200      2210      2220      2230
CTCCCAAGCCCGAGGCGACCCGACGAGGCCCCGAGGGAATAGAGAG
|||||
CTCCCAAGCCCGAGGCGACCCGACGAGGCCCCGAGGGAACCGAAGAAGGAGGTGGAGAGAGAG
12680      12690      12700      12710      12720      12730

2270      2280      2290      2300      2310      2320      2330
TCCATTGCGATTATGTGAACGCGATCCCTAGCACTTATCTGGGACGATCTGCGGAGCCCTTGTGCCTCTTCAGCTA
|||||
TCCATTGCGATTATGTGAACGCGATCCCTAGCACTTATCTGGGACGATCTGCGGAGCC-TGTGCCTCTTCAGCTA
12750      12760      12770      12780      12790      12800      12810

2340      2350      2360      2370      2380      2390      2400      2410
CCACCGCTTGAGAGAGCTTACTCTTGATTGTAAAGGAGGATTGTGGAACCTTCTGGGACGCGAGGGGGGTGGGAAGC
|||||
CCACCGCTTGAGAGAGCTTACTCTTGATTGTAAAGGAGGATTGTGGAACCTTCTGGGACGCGAGGGGGGTGGGAAGT
12820      12830      12840      12850      12860      12870      12880

      12420      12430      12440      12450      12460
CCTCAAAATATTGTTGAATCTCCTACAGTATTGGAGTCAGGAAGCTAAAG
|||||
CCTCAAAATATTGTTGAATCTCCTACAGTATTGGAGTCAGGAAGCTAAAG
12890      12900      12910      12920      12930      X

```

4. KUNZ-158-CL32. SEQ

HIVPV22 Human immunodeficiency virus type 1, isolate PV22.

LOCUS	HIVPV22	8770 bp ss-RNA	VRL	15-JUN-1989
DEFINITION	Human immunodeficiency virus type 1, isolate PV22, complete genome (H9/HTLV-III proviral DNA).			
ACCESSION	K02033			
KEYWORDS	TAR protein; acquired immune deficiency syndrome; complete genome; env gene; gag gene; long terminal repeat; pol gene; polyprotein; provirus; rev gene; reverse transcriptase; tat gene; trans-activator.			
SOURCE	Human immunodeficiency virus type 1 (HIV-1), isolate PV22 (from H9-derived family), proviral DNA.			
ORGANISM	Human immunodeficiency virus type 1 Virodact; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.			
REFERENCE	1 (bases 1 to 8770; revised sequence, personal communication)			
AUTHORS	Muesing, M. A., Smith, D. H., Cabradilla, C. D., Benton, C. V., Lesky, L. A. and Capon, D. J.			
TITLE	Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus			
JOURNAL	Nature 313, 450-458 (1985)			
STANDARD	full staff_review			
REFERENCE	2 (bases 2111 to 2112; revises [1])			
AUTHORS	Muesing, M. A.			
JOURNAL	Unpublished (1987) Whitehead Inst Cambridge, Mass			
STANDARD	full staff_review			
COMMENT	This sequence for a H9/HTLV-III virus was determined from one complete proviral clone [1]. Additionally, several cDNA clones of the viral RNA were sequenced for comparison with the entire proviral sequence. The differences between cDNA and proviral DNA are extensive and are listed in the Sites Table as variations. The authors believe that the variations may be due in part to different strains in the H9/HTLV-III cell line, because it was established by infection with material from several AIDS patients. With the addition of g at 2111, gag cds and pol cds are very close to those of HX22, BRU, and related HIV viruses. For details and other references pertaining to Sites and Features, see the HIV reference entry.			
FEATURES	from	to/span	description	
	289	2337	gag polyprotein precursor	

pept	<	2034	5141	pol polyprotein (NH2-terminus uncertain; AA at 2034)
pept		5088	5804	vif protein
pept		5804	5840	vpr protein
pept		5878	6090	tat protein, exon 2 (first expressed exon)
		8421	8486	tat protein, exon 3 (AA at 8422)
pept		6015	6090	rev protein, exon 2 (first expressed exon)
		8421	8895	rev protein, exon 3 (AA at 8423)
pept		6107	6352	vpu protein
pept		6267	6837	envelope polyprotein
pept		8839	9453	nef protein
pre-m8g		464	9678	genomic mRNA
pre-m8g		464	9678	tat, rev, nef subgenomic mRNA
IVS		753	5522	tat, rev, nef subgenomic mRNA intron 1
IVS		6091	8420	tat cds intron 2
IVS		6091	8420	rev cds intron 2
IVS		6091	8420	tat, rev, nef subgenomic mRNA intron 2
LTR		10	543	5' LTR
LTR		8126	9781	3' LTR
rpt		463	580	R repeat 5' copy
rpt		9581	9678	R repeat 3' copy
binding		386	395	Spl binding site III
binding		397	406	Spl binding site II
binding		408	417	Spl binding site I
binding		645	662	primer (Lys-tRNA) binding site
variant		510	510	a in provirus; g in cDNA [1]
variant		575	575	g in provirus; a in cDNA [1]
revision		2111	2112	gg in [2]; g in [1]
variant		5716	5716	g in provirus; a in cDNA [1]
variant		5992	5992	a in provirus; g in cDNA [1]
variant		6007	6007	c in provirus; t in cDNA [1]
variant		6047	6047	c in provirus; g in cDNA [1]
variant		6051	6051	c in provirus; a in cDNA [1]
variant		6055	6057	agg in provirus; gaa in cDNA [1]
variant		6108	6108	t in provirus; c in cDNA [1]
variant		6120	6120	a in provirus; c in cDNA [1]
variant		6125	6126	gc in provirus; gtaac in cDNA [1]
variant		6136	6136	a in provirus; c in cDNA [1]
variant		6235	6235	t in provirus; a in cDNA [1]
variant		6352	6352	g in provirus; a in cDNA [1]
variant		6760	6760	t in provirus; a in cDNA [1]
variant		7090	7090	c in provirus; t in cDNA [1]
variant		7100	7100	a in provirus; g in cDNA [1]
variant		7134	7135	ca in provirus; ac in cDNA [1]
variant		7193	7194	gt in provirus; aa in cDNA [1]
variant		7199	7199	a in provirus; g in cDNA [1]
variant		7284	7295	aa in provirus; gc in cDNA [1]
variant		7303	7303	a in provirus; c in cDNA [1]
variant		7511	7511	a in provirus [1]; c in cDNA [1]
variant		7533	7533	t in provirus [1]; a in cDNA [1]
variant		7586	7586	c in provirus [1]; t in cDNA [1]
variant		7648	7648	a in provirus [1]; g in cDNA [1]
variant		8139	8139	a in provirus; c in cDNA [1]
variant		8143	8143	t in provirus; c in cDNA [1]
variant		8222	8222	g in provirus; a in cDNA [1]
variant		8269	8269	a in provirus [1]; g in cDNA [1]
variant		8285	8285	g in provirus [1]; t in cDNA [1]
variant		8376	8376	a in provirus [1]; g in cDNA [1]
variant		8381	6381	a in provirus [1]; g in cDNA [1]
variant		8476	8476	a in provirus [1]; g in cDNA [1]
variant		8869	8869	a in provirus [1]; g in cDNA [1]
variant		8978	8979	c in provirus; t in cDNA [1]
variant		8980	8980	a in provirus; c in cDNA [1]
variant		8989	8989	c in provirus [1]; a in cDNA [1]
variant		9031	9031	a in provirus [1]; g in cDNA [1]
variant		9281	9281	t in provirus [1]; g in cDNA [1]

ORIGIN 482 bp upstream of BglII site.

Gaps = 110 Conservative Substitutions = 0

520 530 540 550 560 570 580
TATCAGCACAAGCATAAGAGGTAAAGGTGCAGAAAGAATATGCATTTTTTTTATAAACTTGATATAATACCA
TT
TATCAGCACAAGCATAAGAGGTAAAGGTGCAGAAAGAATATGCATTTTTTTTATAAACTTGATATAATACCA

6750 6760 6770 6780 6790 6800 6810
 590 500 610 620 630 640 650
 AGATAATGATACTACCAAGCTATACGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGT
 |||||
 AGATAATGATACTACCAAGCTATACGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGT
 6820 6830 6840 6850 6860 6870 6880
 660 670 680 690 700 710 720 730
 ATCCCTTCAGCCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAATGTAATAATAAGAC
 |||||
 ATCCCTTCAGCCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAATGTAATAATAAGAC
 6890 6900 6910 6920 6930 6940 6950 6960
 740 750 760 770 780 790 800
 GTTCAATGGAACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATC
 |||||
 GTTCAATGGAACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATC
 6970 6980 6990 7000 7010 7020 7030
 810 820 830 840 850 860 870
 AACTCAACTGCTGTTGAATGGCAGTCTAGCAGAGGAGAGAGGTAGTAATTAGATCTGCCAATTTACAGACAA
 |||||
 AACTCAACTGCTGTTGAATGGCAGTCTAGCAGAGGAGAGAGGTAGTAATTAGATCTGCCAATTTACAGACAA
 7040 7050 7060 7070 7080 7090 7100
 880 890 900 910 920 930 940
 TGCTAAAACCATTAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAG
 |||||
 TGCTAAAACCATTAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAG
 7110 7120 7130 7140 7150 7160 7170
 950 960 970 980 990 1000 1010
 AAAAAGTATCCGTATCCAGAGAGGACCAAGGGAGAGCATTGTGTTACAATAGGAAAAATAGGAAATATGAGACA
 |||||
 AAAAAGTATCCGTATCCAGAGAGGACCAAGGGAGAGCATTGTGTTACAATAGGAAAAATAGGAAATATGAGACA
 7180 7190 7200 7210 7220 7230 7240
 1020 1030 1040 1050 1060 1070 1080 1090
 AGCAGATTGTAACATTAGTAGAGCAAAATGCAATGCCACTTTAAACAGATAGCTAGCAAAATTAAGAGAACA
 |||||
 AGCAGATTGTAACATTAGTAGAGCAAAATGCAATGCCACTTTAAACAGATAGCTAGCAAAATTAAGAGAACA
 7250 7260 7270 7280 7290 7300 7310 7320
 1100 1110 1120 1130 1140 1150 1160
 ATTTGGAAATTAATAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTT
 |||||
 ATTTGGAAATTAATAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTT
 7330 7340 7350 7360 7370 7380 7390
 1170 1180 1190 1200 1210 1220 1230
 TAATTGTGGAGGGGAATTTTCTACTGTAATTCACACAACCTGTTTAATAGTACTTGGTTTAATAGTACTTG
 |||||
 TAATTGTGGAGGGGAATTTTCTACTGTAATTCACACAACCTGTTTAATAGTACTTGGTTTAATAGTACTTG
 7400 7410 7420 7430 7440 7450 7460
 1240 1250 1260 1270 1280 1290 1300
 GAGTACTGAAGGGTCAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
 |||||
 GAGTACTGAAGGGTCAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
 7470 7480 7490 7500 7510 7520 7530
 1310 1320 1330 1340 1350 1360 1370
 AAACATGTTGGGGGGAAGTAGCAATGCAATGATGCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAAA
 |||||
 AAACATGTTGGGGGGAAGTAGCAATGCAATGATGCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAAA

7540	7550	7560	7570	7580	7590	7600	
1380	1390	1400	1410	1420	1430	1440	1450
TATTACAGGGCTGCTATTAAACAGAGATGGTGGTAATAACAACAATGGGTCCGAGATCTTCAGACCTGGAGG							
TATTACAGGGCTGCTATTAAACAGAGATGGTGGTAATAACAACAATGGGTCCGAGATCTTCAGACCTGGAGG							
7610	7620	7630	7640	7650	7660	7670	7680
1460	1470	1480	1490	1500	1510	1520	
AGGAGATATGAGGGACAATTGGAGCAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGT							
AGGAGATATGAGGGACAATTGGAGCAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGT							
7690	7700	7710	7720	7730	7740	7750	
1530	1540	1550	1560	1570	1580	1590	
AECACCCACCAAGGCCAAGAGAGAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTTCCT							
AECACCCACCAAGGCCAAGAGAGAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTTCCT							
7760	7770	7780	7790	7800	7810	7820	
1600	1610	1620	1630	1640	1650	1660	
TGGGTTCTTGGGAGCAGCAGGAACCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATT							
TGGGTTCTTGGGAGCAGCAGGAACCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATT							
7830	7840	7850	7860	7870	7880	7890	
1670	1680	1690	1700	1710	1720	1730	
ATTGTTCTGCTATAGTGCAGCAGCAGAACCAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACT							
ATTGTTCTGCTATAGTGCAGCAGCAGAACCAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACT							
7900	7910	7920	7930	7940	7950	7960	
1740	1750	1760	1770	1780	1790	1800	1810
CACAGTCTTGGGECATCAAGCAGCTCCAGGCCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCT							
CACAGTCTTGGGECATCAAGCAGCTCCAGGCCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCT							
7970	7980	7990	8000	8010	8020	8030	8040
1820	1830	1840	1850	1860	1870	1880	
CCTGGGGATTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAA							
CCTGGGGATTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAA							
8050	8060	8070	8080	8090	8100	8110	
1890	1900	1910	1920	1930	1940	1950	
TAAATCTCTGGACAGATTTGGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAG							
TAAATCTCTGGACAGATTTGGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAG							
8120	8130	8140	8150	8160	8170	8180	
1960	1970	1980	1990	2000	2010	2020	
CTTAATACACTCTTAATTGAAGAAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGA							
CTTAATACACTCTTAATTGAAGAAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGA							
8190	8200	8210	8220	8230	8240	8250	
2030	2040	2050	2060	2070	2080	2090	
TAAATGGGCAAGTTTGTGGGAATTGTTTAAACATAACAAATTGGCTGTGGTATATAAAAAATTATTCATAATGAT							
TAAATGGGCAAGTTTGTGGGAATTGTTTAAACATAACAAATTGGCTGTGGTATATAAAAAATTATTCATAATGAT							
8260	8270	8280	8290	8300	8310	8320	
2100	2110	2120	2130	2140	2150	2160	2170
AGTAGGAGTCTTGGTATGTTTAAACATAGTTTTTCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATA							
AGTAGGAGTCTTGGTATGTTTAAACATAGTTTTTCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATA							

8330	8340	8350	8360	8370	8380	8390	8400
2180	2190	2200	2210	2220	2230	2240	
TTCACCATTTATCGTTTTCAGACCCACCTCCCAACCCCGAGGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGA							
TTCACCATTTATCGTTTTCAGACCCACCTCCCAACCCCGAGGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGA							
8410	8420	8430	8440	8450	8460	8470	
2250	2260	2270	2280	2290	2300	2310	
AGGATGGAGAGAGAGACAGAGACGATCCATTGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCT							
AGATGGAGAGAGAGACAGAGACGATCCATTGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCT							
8480	8490	8500	8510	8520	8530	8540	
2320	2330	2340	2350	2360	2370	2380	
GCGGAGCCCTGATGCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAAC							
GCGGAGCCCTGATGCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAAC							
8550	8560	8570	8580	8590	8600	8610	
2390	2400	2410	2420	2430	2440	2450	
TTCATGGAGGCGAGGGGATGGGAAGCCCTCAAATATTGGTGGGAATCTCCTACAGTATTGGAGTCAGGAACTAA							
TTCATGGAGGCGAGGGGATGGGAAGCCCTCAAATATTGGTGGGAATCTCCTACAGTATTGGAGTCAGGAGCTAA							
8620	8630	8640	8650	8660	8670	8680	

X
AG
||
AG
8690

5. KUNZ-158-CL32L SEQ

HIVBH102 Human immunodeficiency virus type 1, isolate BH10,

LOCUS	HIVBH102	8932 bp ss-RNA	VRL	15-JUN-1989
DEFINITION	Human immunodeficiency virus type 1, isolate BH10, genome.			
ACCESSION	M15584 KC2010 KC2008 KC2009			
KEYWORDS	TAD region; acquired immune deficiency syndrome; env gene; gag gene; long terminal repeat; pol gene; polyprotein; provirus; reverse transcriptase; trans-activator.			
SEGMENT	2 of 2			
SOURCE	Human immunodeficiency virus type 1 (HIV-1) proviral DNA clone BH10.			
ORGANISM	Human immunodeficiency virus type 1 Virusidae; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.			
REFERENCE	1 (bases 1 to 8932)			
AUTHORS	Ratner,L. , Haseltine,W. , Patarca,R. , Livak,K. J. , Starcich,B. , Josephs,S. F. , Doran,E. R. , Rafalski,J. A. , Whitehorn,E. A. , Baumelster,K. , Ivanoff,L. , Petteway,S. R. Jr. , Pearson,M. L. , Lautenberger,J. A. , Papes,T. S. , Ghayeb,J. , Chang,N. T. , Gallo,R. C. and Wong-Staal,F.			
TITLE	Complete nucleotide sequence of the AIDS virus, HTLV-III			
JOURNAL	Nature 313, 277-284 (1985)			
STANDARD	full staff_review			
COMMENT	The BH10 sequence differs from BH8 and BH5 by 0.9% in the coding regions and 1.8% in the noncoding regions, and the authors of [1] believe that these are stable variants. The HTLV-III genome encodes at least seven proteins: gag, pol, env, tat, trs, 27K antigen and the sor 23K product. The 3' ORF (positions 8153-8773) is truncated in BH10 (stop codon at positions 8522-8524), but reads through in BH8 and other sequences to yield what is now called the 27K antigen. The mechanism for pol gene translation has not been elucidated; a			

gag-pol fusion protein is possible; splicing or frameshift have not been ruled out. The viral protease would be determined by the region in question.

The Tat protein (trans-activator protein, approximately 14 kb) is an effector of an autostimulatory pathway through interaction with a positive control element, the trans-activating responsive sequence, TAR. tat seems to be a transcriptional control molecule in HTLV-I, but is both that and a post-transcriptional regulatory molecule in HTLV-III. Deletion mutants in the tat gene are incapable of prolific replication and exhibit no cytopathic effects in T4 cell lines.

In addition to the 7.4 kb genomic mRNA, subgenomic mRNAs of 7.4, 5.5, 5.0, 4.3, 2.0 and 1.8 kb have been detected.

FEATURES	from	to/span	description
pept	112	1650	gag polyprotein precursor
pept	1407	4454	pol polyprotein (NH2-terminus uncertain; AA at 1407)
pept	4395	4977	vif protein
pept	4917	5153	vpr protein
pept	5135	5403	tat protein, exon 2 (first expressed exon)
	7734	7779	tat protein, exon 3 (AA at 7735)
pept	5325	5403	rev protein, exon 2 (first expressed exon)
	7734	8002	rev protein, exon 3 (AA at 7736)
pept	5420	5665	vpu protein
pept	5550	8150	envelope polyprotein
pept	5152	8523	nef protein, exon 3 (first expressed exon; premature termination)
pre-msg	1	8932	genomic mRNA
pre-msg	1	8932	tat, rev, nef subgenomic mRNA
IVS	66	5135	tat, rev, nef subgenomic mRNA intron 1
IVS	5404	7733	tat cds intron 2
IVS	5404	7733	rev cds intron 2
IVS	5404	7733	tat, rev, nef subgenomic mRNA intron 2
LTR	8441	8932	3' LTR
rpt	8294	8932	R repeat 3' copy
BASE COUNT	3227 a	1582 c	2147 g 1976 t
ORIGIN	About 120 bp downstream from HIVBH101.		

Initial Score = 1872 Optimized Score = 2186 Significance = 0.00
 Residue Identity = 89% Matches = 2255 Mismatches = 157
 Gaps = 110 Conservative Substitutions = 0

```

X          10          20          30          40          50          60
AAGAG-CAG---AAGACAGTGGCAATGAGAGTGAABGAGAA-ATATCAGCACTTGTGGAGA-TGGGGGTGGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AATGACAGGTTAATTGATAGACTAATAGAAAG-AGCAGAAGACAGTGGCAAT----GAGAGTGAAGGAGAA
5530      5540      5550      5560      5570      5580      5590

      70      80      90          100          110          120
AATGGGGCAG-CATGCTCCTTGGGATATTGATG-----AT-CT-GTAGTGCTACAGAAAAATTGT-GGGT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ATATCAGCACTTGTGAGATGGGGGTGGAGATGGGGGCACCATGCTCCTTGGGATGTTGATGATCTGTAGTG
5600      5610      5620      5630      5640      5650      5660

      130      140      150          160          170          180
CACAG-TCTATTATGGGGTAC---CT---GTGTGGAA-----GGAAGCAA-CCACCA-CTCTATTTTGTG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TACAGAAATTTGTGCTCAGASTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCACTCTATTTTG
5670      5680      5690      5700      5710      5720      5730      5740

      190      200      210          220          230          240
CATCAGATGCTAAGGCATATGATA---CAGAGG-TACATA-----AT--GTTTGGGGCCACACATGCCTG--T
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
-----TGC---ATCAGATGCTAAGGCATATGATACAGAGGTACATAATGTTTGGGGC-CACA--CATGCCCT
5750      5760      5770      5780      5790

```

[illegible]

1020 1030 1040 1050 1060 1070 1080 1090
AGCACATTGTATGATTAAGTACAGCAAAATGCAATGCCACTTTAAACAGATAGCTAGCAAATTAAGAGAACAA
|||||
AGCACATTGTATGATTAAGTACAGCAAAATGGAATAACACTTTAAACAGATAGATAGCAAATTAAGAGAACAA
6570 6580 6590 6600 6610 6620 6630

1100 1110 1120 1130 1140 1150 1160
ATTTTGAAATATTAAGACAAATATCTTTAAGCAATCCTCAGGAGGGGGACCCAGAAATTGTAACGCACAGTTT
|||||
ATTTTGAAATATTAAGACAAATATCTTTAAGCAATCCTCAGGAGGGGGACCCAGAAATTGTAACGCACAGTTT
6640 6650 6660 6670 6680 6690 6700

1170 1180 1190 1200 1210 1220 1230
TAATTTGTGAGGAGGCAATTTTCTACTGTAAATTCACACAACTGTTTAAATAGTACTTGGTTTAAATAGTACTTG
|||||
TAATTTGTGAGGAGGCAATTTTCTACTGTAAATTCACACAACTGTTTAAATAGTACTTGGTTTAAATAGTACTTG
6710 6720 6730 6740 6750 6760 6770

1240 1250 1260 1270 1280 1290 1300
GAGTACTGAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
|||||
GAGTACTGAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
6780 6790 6800 6810 6820 6830 6840 6850

1310 1320 1330 1340 1350 1360 1370
AAACATGTTGAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
|||||
AAACATGTTGAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
6860 6870 6880 6890 6900 6910 6920

1380 1390 1400 1410 1420 1430 1440 1450
TATTTACAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
|||||
TATTTACAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
6930 6940 6950 6960 6970 6980 6990

1460 1470 1480 1490 1500 1510 1520
AGGATATATGAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
|||||
AGGATATATGAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
7000 7010 7020 7030 7040 7050 7060

1530 1540 1550 1560 1570 1580 1590
AGCACATGAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
|||||
AGCACATGAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
7070 7080 7090 7100 7110 7120 7130

1600 1610 1620 1630 1640 1650 1660
TGGGTTCTTGGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
|||||
TGGGTTCTTGGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
7140 7150 7160 7170 7180 7190 7200 7210

1670 1680 1690 1700 1710 1720 1730
ATTTGCTGAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
|||||
ATTTGCTGAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
7220 7230 7240 7250 7260 7270 7280

1740 1750 1760 1770 1780 1790 1800 1810
GAGATATGAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
|||||
GAGATATGAGGAGGCAATTAAGCAATGAGGAGGAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
7290 7300 7310 7320 7330 7340 7350

[illegible]

[illegible]

```

1850      1860      1870      1880      1890      1900      1910
CTCTGGAGACASATTTTAAATTAAGATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAA
|||||
1850      1860      1870      1880      1890      1900      1910
CTCTGGAGACASATTTTAAATTAAGATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAA
|||||
2330      2340      2350      2360      2370      2380      2390      2400

1960      1970      1980      1990      2000      2010      2020      2030
TACATTCCTTATATTCAGAAATTTTAAAACCAAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATAAAT
|||||
1960      1970      1980      1990      2000      2010      2020      2030
TACATTCCTTATATTCAGAAATTTTAAAACCAAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATAAAT
|||||
2410      2420      2430      2440      2450      2460      2470

2040      2050      2060      2070      2080      2090      2100
GGGCAAGTTTGTGGAAATTGCTTAAACAATACAAATTGGCTGTGGTATATAAAAATATTTCATAATGATAGTAG
|||||
2040      2050      2060      2070      2080      2090      2100
GGGCAAGTTTGTGGAAATTGCTTAAACAATACAAATTGGCTGTGGTATATAAAAATATTTCATAATGATAGTAG
|||||
2480      2490      2500      2510      2520      2530      2540

2110      2120      2130      2140      2150      2160      2170
GAGGCTTGGTAAATTTAAGAAATATTTTGGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCAC
|||||
2110      2120      2130      2140      2150      2160      2170
GAGGCTTGGTAAATTTAAGAAATATTTTGGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCAC
|||||
2550      2560      2570      2580      2590      2600      2610

2180      2190      2200      2210      2220      2230      2240
CATTATCGTTTCAGACCCAGCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTG
|||||
2180      2190      2200      2210      2220      2230      2240
CATTATCGTTTCAGACCCAGCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTG
|||||
2620      2630      2640      2650      2660      2670      2680

2250      2260      2270      2280      2290      2300      2310
GAGAGAGAGACAGACAGATCCATTCGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCTGCGGA
|||||
2250      2260      2270      2280      2290      2300      2310
GAGAGAGAGACAGACAGATCCATTCGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCTGCGGA
|||||
2690      2700      2710      2720      2730      2740      2750      2760

2320      2330      2340      2350      2360      2370      2380      2390
GCCTTGTCCTCTTCAGCTACCAACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACCTTCTG
|||||
2320      2330      2340      2350      2360      2370      2380      2390
GCC-TGTCCTCTTCAGCTACCAACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACCTTCTG
|||||
2770      2780      2790      2800      2810      2820      2830

2400      2410      2420      2430      2440      2450      X
GGACGCAGGGGGTGGGAAGCCCTCAAAATATTGGTGGAAATCTCCTACAGTATTGGAGTCAGGAACTAAAG
|||||
2400      2410      2420      2430      2440      2450      X
GGACGCAGGGGGTGGGAAGCCCTCAAAATATTGGTGGAAATCTCCTACAAATATTGGAGTCAGGAGCTAAAG
|||||
2840      2850      2860      2870      2880      2890      2900

```

7. KUNZ-158-CL32. SEQ

HIVNL43 Human immunodeficiency virus type 1, NY5/BRU (LAV-

LOCUS	HIVNL43	8708 bp ss-RNA	VRL	15-JUN-1989
DEFINITION	Human immunodeficiency virus type 1, NY5/BRU (LAV-1) recombinant clone pNL4-3.			
ACCESSION	M19921			
KEYWORDS	.			
SOURCE	Human immunodeficiency virus type 1 (HIV-1), NY5/BRU (LAV-1) recombinant clone pNL4-3.			
ORGANISM	Human immunodeficiency virus type 1 Virus; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.			
REFERENCE	1. (bases 1 to 8708)			
AUTHORS	Buck, A. C., Buckler-White, A. J., Willey, R. L. and McCoy, J.			
JOURNAL	Univ. of Calif. (1988).			
STANDARD	HIV-1 standard			


```

new1:100  1715  1700  forced in [31] ctcaca in [1]
new1:100  1716  1701  c in [31] a in [1]
stop:1    1697  1697  mRNA polyadenylation signal
BASE: 31007 1697 a 1700 c 1706 g 2166 t
ORIGIN: 2166 bp of cDNA of 2175 bp

```

Initial score	=	1721	Optimized Score	=	2181	Significance	=	0.00
Residual identity	=	15.1	Aligns	=	2238	Mismatches	=	187
Gaps	=	0	Comparative Substitutions	=			=	0

[illegible]

6780 6790 6800 6810 6820 6830 6840 6850

6860 6870 6880 6890 6900 6910 6920

6930 6940 6950 6960 6970 6980 6990

7000 7010 7020 7030 7040 7050 7060

7070 7080 7090 7100 7110 7120 7130

7140 7150 7160 7170 7180 7190 7200 7210

7220 7230 7240 7250 7260 7270 7280

7290 7300 7310 7320 7330 7340 7350

7360 7370 7380 7390 7400 7410 7420

7430 7440 7450 7460 7470 7480 7490

7500 7510 7520 7530 7540 7550 7560 7570

7580 7590 7600 7610 7620 7630 7640 7650

TTTGTAGACCGACCTTCGATCTCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGA

8. KUNZ-158-0130, FBI

LOCUS HTLV28000 9718 bp ss-RNA VRL 25-SEP-1987
DEFINITION Human immunodeficiency virus type 1 (HXB2), complete genome;
HIV-1/HTLV-III/LAV reference genome.
ACCESSION F01425
KEYWORDS TAR protein; acquired immune deficiency syndrome; complete genome;
env gene; gag gene; long terminal repeat; pol gene; polyprotein;
provirus; reverse transcriptase; trans-activator.
SOURCE HTLV-III/LAV (isolate HXB2) proviral DNA.
ORGANISM Human immunodeficiency virus type 1
Virus; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.
REFERENCE 1 (bases 1 to 653)
AUTHORS Starcich,B., Ratner,L., Josephs,S.F., Okamoto,T., Gallo,R.C. and
Wong-Staal,F.
TITLE Characterization of long terminal repeat sequences of HTLV-III
JOURNAL Science 227: 538-540 (1985)
STANDARD full staff_review
REFERENCE 2 (bases 458 to 674; 9577 to 9718)
AUTHORS Ratner,L., Haseltine,W., Patarca,R., Livak,K.J., Starcich,B.,
Josephs,S.J., Doran,E.R., Rafalski,J.A., Whitehorn,E.A.,
Baumeister,K., Ivanoff,L., Petteway,S.R. Jr., Pearson,M.L.,
Lautenberger,J.A., Papas,T.S., Ghayeb,J., Chang,N.T.,
Gallo,R.C. and Wong-Staal,F.
TITLE Complete nucleotide sequence of the AIDS virus, HTLV-III
JOURNAL Nature 313: 277-284 (1985)
STANDARD full staff_review
REFERENCE 3 (bases 8761 to 9060)
AUTHORS Fisher,A.G., Ratner,L., Mitsuya,H., Marselle,L.M., Harper,M.E.,
Broder,S., Gallo,R.C. and Wong-Staal,F.
TITLE Infectious mutants of HTLV-III with changes in the 3' region and
markedly reduced cytopathic effects
JOURNAL Science 233: 655-658 (1986)
STANDARD full staff_review
REFERENCE 4 (bases 1 to 9635)
AUTHORS Ratner,L., Fisher,A., Jagodzinski,L.L., Mitsuya,H., Liou,R.-S.,
Gallo,R.C. and Wong-Staal,F.
TITLE Complete nucleotide sequences of functional clones of the AIDS
virus
JOURNAL AIDS Res. Hum. Retroviruses 3: 57-69 (1987)
STANDARD full staff_review
REFERENCE 5 (bases: tat mRNA and other transcript boundaries)

AUTHORS Anya, S.K. , Sud, C. , Josephs, S.F. and Wong-Staal, F.
 TITLE Trans-activator gene of human T-lymphotropic virus type III (HTLV-III)
 JOURNAL Science 228, 69-73 (1985)
 STANDARD full staff_review

REFERENCE 6 (sites; tat mRNA)
 AUTHORS Sodroski, J. , Paterce, R. , Rosen, C. , Wong-Staal, F. and Haseltine, W.
 TITLE Location of the trans-activating region on the genome of human T-cell lymphotropic virus type III
 JOURNAL Science 228, 74-77 (1985)
 STANDARD full staff_review

REFERENCE 7 (sites; mRNA splice sites)
 AUTHORS Rabson, A. E. , Daugherty, D. F. , Venkatesan, S. , Boulukos, K. E. , Benn, S. I. , Folks, T. M. , Feorino, P. and Martin, M.
 TITLE Transcription of novel open reading frames of AIDS retrovirus during infection of lymphocytes
 JOURNAL Science 229, 1388-1390 (1985)
 STANDARD full staff_review

REFERENCE 8 (sites; 27K antigen cds)
 AUTHORS Allen, J. S. , Coligan, J. E. , Lee, T. -H. , McLane, M. F. , Kanki, P. J. , Gronpman, J. E. and Essex, M.
 TITLE A new HTLV-III/LAV encoded antigen detected by antibodies from AIDS patients
 JOURNAL Science 230, 810-813 (1985)
 STANDARD full staff_review

REFERENCE 9 (sites; gp160 and gp120 coding sequences)
 AUTHORS Allen, J. S. , Coligan, J. E. , Barin, F. , McLane, M. F. , Sodroski, J. G. , Rosen, C. A. , Haseltine, W. A. , Lee, T. H. and Essex, M.
 TITLE Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III
 JOURNAL Science 228, 1091-1094 (1985)
 STANDARD full staff_review

REFERENCE 10 (sites; regulatory sequences in the LTR)
 AUTHORS Rosen, C. A. , Sodroski, J. G. and Haseltine, W. A.
 TITLE The location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat
 JOURNAL Cell 41, 813-823 (1985)
 STANDARD full staff_review

REFERENCE 11 (review; bases 1 to 9718)
 AUTHORS Van Beveren, C. , Coffin, J. and Hughes, S.
 TITLE Appendix B: HTLV-3 genome
 JOURNAL (in) Weiss, R. , Teich, N. , Varmus, H. and Coffin, J. (Eds.); RNA Tumor Viruses, Second Edition, 2: 1102-1123 and 1147-1148; Cold Spring Harbor Laboratory, New York (1985)
 STANDARD full staff_review

REFERENCE 12 (sites; trans-activator function and TAR sequence)
 AUTHORS Rosen, C. A. , Sodroski, J. G. , Goh, W. C. , Dayton, A. I. , Lippke, J. and Haseltine, W. A.
 TITLE Post-transcriptional regulation accounts for the trans-activation of the human T-lymphotropic virus type III
 JOURNAL Nature 319, 555-558 (1986)
 STANDARD full staff_review

REFERENCE 13 (sites; pol coding sequence)
 AUTHORS di Marzo Veronese, F. , Copeland, T. D. , DeVico, A. L. , Rahman, R. , Grosslan, S. , Gallo, R. C. and Sarngadharan, M. G.
 TITLE Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV
 JOURNAL Science 231, 1289-1291 (1986)
 STANDARD full staff_review

REFERENCE 14 (sites; 23K src gene product)
 AUTHORS Han, N. G. , Frenchini, G. , Wong-Staal, F. , DuBois, G. C. , Robey, W. G. , Lautenberger, J. A. and Papas, T. S.
 TITLE Identification of HTLV-III/LAV src gene product and detection of antibodies in human sera
 JOURNAL Science 231, 1553-1555 (1986)
 STANDARD full staff_review

REFERENCE 15 (sites; pol NH2-terminal region)
AUTHORS Kramer, R. A. , Schaber, M. D. , Skalka, A. M. , Ganguly, K. ,
Wong-Staal, F. and Reddy, E. P.
TITLE HTLV-III gag protein is processed in yeast cells by the virus
pol-protease
JOURNAL Science 231: 1580-1584 (1986)
STANDARD full staff_review

REFERENCE 16 (sites; sor 23K protein)
AUTHORS Lse, T.-H. , Coligan, J. E. , Allan, J. S. , McLane, M. F. , Groopman, J. E. and
Essex, M.
TITLE A new HTLV-III/LAV protein encoded by a gene found in cytopathic
retroviruses
JOURNAL Science 231: 1546-1549 (1986)
STANDARD full staff_review

REFERENCE 17 (sites; sor 23K protein)
AUTHORS Sodroski, J. , Goh, W. C. , Rosen, C. , Tartar, A. , Portetelle, D. ,
Rumy, A. and Haseltine, W.
TITLE Replicative and cytopathic potential of HTLV-III/LAV with sor gene
deletions
JOURNAL Science 231: 1549-1553 (1986)
STANDARD full staff_review

REFERENCE 18 (sites; Spi binding sites in the promoter region)
AUTHORS Jones, K. A. , Kedonaga, J. T. , Luciw, P. A. and Tjian, R.
TITLE Activation of the AIDS retrovirus promoter by the cellular
transcription factor, Spi
JOURNAL Science 232: 755-759 (1986)
STANDARD full staff_review

REFERENCE 19 (sites; acceptor and donor splice sites for tat and 27K)
AUTHORS Arya, S. K. and Gallo, R. C.
TITLE Three novel genes of human T-lymphotropic virus type III: Immune
reactivity of their products with sera from acquired immune
deficiency syndrome patients
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83: 2209-2213 (1986)
STANDARD full staff_review

REFERENCE 20 (sites; deletion mutants in the tat gene)
AUTHORS Dayton, A. I. , Sodroski, J. G. , Rosen, C. A. , Goh, W. C. and Haseltine, W. A.
TITLE The trans-activator gene of the human T cell lymphotropic virus
type III is required for replication
JOURNAL Cell 44: 541-547 (1986)
STANDARD full staff_review

REFERENCE 21 (sites; env gene conserved/variable regions; separate entries)
AUTHORS Willey, R. , Rutledge, R. A. , Dias, S. , Folks, T. , Theodore, T. ,
Buckler, C. E. and Martin, M. A.
TITLE Identification of conserved and divergent domains within the
envelope gene of the acquired immunodeficiency syndrome virus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83: 5038-5042 (1986)
STANDARD full staff_review

REFERENCE 22 (sites; trs cds boundaries)
AUTHORS Sodroski, J. , Goh, W. C. , Rosen, C. , Dayton, A. I. , Terwilliger, E. and
Haseltine, W.
TITLE A second post-transcriptional trans-activator gene required for
HTLV-III replication
JOURNAL Nature 321: 412-417 (1986)
STANDARD full staff_review

REFERENCE 23 (sites; trs cds boundaries)
AUTHORS Feinberg, M. B. , Jarret, R. F. , Aldovini, A. , Gallo, R. C. and
Wong-Staal, F.
TITLE HTLV-III expression and production involve complex regulation at
the levels of splicing and translation of viral RNA
JOURNAL Cell 45: 607-617 (1986)
STANDARD full staff_review

REFERENCE 24 (sites; env gene conserved/variable regions; separate entries)
AUTHORS Starcich, B. R. , Hahn, B. H. , Shaw, G. M. , McNeely, P. D. , Modrow, S. ,
Wolfe, H. , Parks, E. S. , Parks, W. P. , Josephs, S. F. , Gallo, R. C. and
Wong-Staal, F.
TITLE Identification and characterization of conserved and variable

regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS

JOURNAL Cell 45, 637-648 (1986)
STANDARD full staff_review

REFERENCE 25 (sites; tar or transactivator target)
AUTHORS Wright,C.M. , Felber,B.K. , Paskalis,H. and Pavlakis,G.N.
TITLE Expression and characterization of the trans-activator of HTLV-III/LAV virus

JOURNAL Science 234, 998-992 (1986)
STANDARD full staff_review

REFERENCE 26 (sites; 3' orf mutations)
AUTHORS Terwilliger,E. , Sodroski,J.G. , Rosen,C.A. and Haseltine,W.A.
TITLE Effects of mutations within the 3' orf open reading frame region of human T-cell lymphotropic virus type III (HTLV-III/LAV) on replication and cytopathogenicity

JOURNAL J. Virol. 60, 754-760 (1986)
STANDARD full staff_review

REFERENCE 27 (sites; pol p34 terminus)
AUTHORS Lightfoote,M.M. , Coligan,J.E. , Folks,T.M. , Fauci,A.S. , Martin,M.A. and Vankatesan,S.
TITLE Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus

JOURNAL J. Virol. 60, 771-775 (1986)
STANDARD full staff_review

REFERENCE 28 (sites; promoter; TAR; tat-III mutants)
AUTHORS Muesing,M.A. , Smith,D.H. and Capon,D.J.
TITLE Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein

JOURNAL Cell 48, 591-701 (1987)
STANDARD full staff_review

REFERENCE 29 (sites; envelope protein; epitopes)
AUTHORS Modrow,S. , Hahn,B.H. , Shaw,G.M. , Gallo,R.C. , Wong-Staal,F. and Wolf,H.
TITLE Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: Prediction of antigenic epitopes in conserved and variable regions

JOURNAL J. Virol. 61, 570-578 (1987)
STANDARD full staff_review

REFERENCE 30 (sites; trs/art protein)
AUTHORS Goh,W.C. , Sodroski,J.G. , Rosen,C.A. and Haseltine,W.A.
TITLE Expression of the art gene protein of human T-lymphotropic virus type III (HTLV-III/LAV) in bacteria

JOURNAL J. Virol. 61, 633-637 (1987)
STANDARD full staff_review

REFERENCE 31 (sites; inducible enhancer element)
AUTHORS Nabel,G. and Baltimore,D.
TITLE An inducible transcription factor activates expression of human immunodeficiency virus in T cells

JOURNAL Nature 326, 711-713 (1987)
STANDARD full staff_review

REFERENCE 32 (bases 5611 to 5611; revises [4])
AUTHORS Rainer,L.
JOURNAL Unpublished (1987) Washington U Med School, St. Louis, MO
STANDARD full staff_review

REFERENCE 33 (sites; long terminal repeat)
AUTHORS Petarca,R. , Heath,C. , Goldenberg,G.J. , Rosen,C.A. , Sodroski,J.G. , Haseltine,W.A. and Hansen,U.M.
TITLE Transcription directed by the HIV long terminal repeat in vitro

JOURNAL AIDS Res. Hum. Retroviruses 3, 41-55 (1987)
STANDARD full staff_review

REFERENCE 34 (sites; R orf)
AUTHORS Wong-Staal,F. , Chanda,P.K. and Ghayeb,J.
TITLE Human immunodeficiency virus: the eighth gene

JOURNAL AIDS Res. Hum. Retroviruses 3, 33-39 (1987)
STANDARD full staff_review

REFERENCE 35 (sites/ sor)

AUTHORS Fisher, A. G., Ensoli, B., Ivanoff, L., Chamberlain, M., Petteway, S., Ratner, L., Gallo, R. C. and Wong-Staal, F.

TITLE The sor gene of hiv-1 is required for efficient virus transmission in vitro

JOURNAL Science 237, 888-893 (1987)

STANDARD full staff_review

COMMENT Sequence for [31] kindly provided in computer-readable form by L. Ratner, 19-AUG-1986.
The HXB2 sequence is being used as a reference genome for all the HIV entries because it has been derived from a demonstrably infectious clone. Hence not all of the "sites" references above were concerned with this isolate.

FEATURES	from	to/span	description
pept	789	2291	gag polyprotein
pept	/ 2357	5095	pol polyprotein (NH2-terminus uncertain; AA at 2357)
pept	5040	5618	sor 23K protein
pept	5558	5794	R (ORF) protein
pept	5830	6044	tat protein, exon 2 (first expressed exon)
	6378	6423	tat protein, exon 3
pept	5858	6044	trs protein, exon 2 (first expressed exon)
	6378	6652	trs protein, exon 3
pept	6224	6794	envelope polyprotein
pept	6796	9167	27K protein (premature termination)
mRNA	455	9635	HXB2 genomic mRNA
pre-msg	455	9635	tat, trs, 27K subgenomic mRNA
IVS	6045	8377	tat intron 1
IVS	6045	8377	trs intron 2
IVS	6045	8377	27K mRNA intron 2
IVS	743	5776	tat, trs, 27K mRNA intron 1
IVS	6045	8377	tat, trs intron 2
LTR	1	634	5' LTR
LTR	9085	9718	3' LTR
rpt	454	551	R repeat 5' copy
rpt	9539	9635	R repeat 3' copy
binding	377	396	Spl binding site III
binding	388	397	Spl binding site II
binding	358	408	Spl binding site I
binding	636	653	primer (Lys-tRNA) binding site
revision	5611	5611	g in [32]; a in [4]
signal	9611	9616	HXB2 mRNA polyadenylation signal
BASE COUNT	3411 a	1773 c	2370 g 2164 t
ORIGIN	435 bp upstream of PvuII site; 5' end of proviral genome.		

Initial Score = 1858 Optimized Score = 2174 Significance = 0.00

Residue Identity = 88% Matches = 2243 Mismatches = 169

Gaps = 110 Conservative Substitutions = 0

```

X          10          20          30          40          50          60
AAGAG-CAG---AAGACAGTGGCAATGAGAGTGAAGGAGAA-ATATCAGCACTTGTGGAGA-TGGGGGTGGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AATAGACAGGTTAATTGATAGACTAATAGAAAG-AGCAGAAGACAGTGGCAAT----GAGAGTGAAGGAGAA
X 6180          6190          6200          6210          6220          6230          6240

          70          80          90          100          110          120
AATGGGGCAC-CATGCTCCTTGGGATATTGATG-----AT-CT-GTAGTGCTACAGAAAAATTGT-GGGT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ATA)CAGCACTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTTGATGATCTGTAGTGC
          6250          6260          6270          6280          6290          6300          6310

          130          140          150          160          170          180
CAGAG-TCATATTATGGGGTAC---CT---GTGTGGAA-----GGAAGCAA-CCACCA-CTCTATTTTGTG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TACAGAAAAATTTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCACCACTCTATTTTG
          6320          6330          6340          6350          6360          6370          6380

```


[illegible]

950 960 970 980 990 1000 1010
AAAAAGTATCCGTATCCAGAGAGGACCAGGGAGAGCATTGTGTTACAATAGGAAAAATAGGAAATATGAGACA
|||||
AAAAAGAATCCGTATCCAGAGAGGACCAGGGAGAGCATTGTGTTACAATAGGAAAAATAGGAAATATGAGACA
7140 7150 7160 7170 7180 7190 7200

1020 1030 1040 1050 1060 1070 1080 1090
AGCACATTGTAAACATTAGTAGAGCAAAATGCAATGCCACTTTAAAACAGATAGCTAGCAAATTAAGAGACA
|||||
AGCACATTGTAAACATTAGTAGAGCAAAATGCAATTAACACTTTAAAACAGATAGCTAGCAAATTAAGAGACA
7210 7220 7230 7240 7250 7260 7270

1100 1110 1120 1130 1140 1150 1160
ATTTCGAAATAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTT
|||
ATTTCGAAATAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTT
7280 7290 7300 7310 7320 7330 7340 7350

1170 1180 1190 1200 1210 1220 1230
TAATTGTGAGGGGAATTTTCTACTGTAATTCAACACAACACTGTTTAAATAGTACTTGGTTTAAATAGTACTTG
|||||
TAATTGTGAGGGGAATTTTCTACTGTAATTCAACACAACACTGTTTAAATAGTACTTGGTTTAAATAGTACTTG
7360 7370 7380 7390 7400 7410 7420

1240 1250 1260 1270 1280 1290 1300
GAGTACTGAGGGGTCAAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTAT
|||||
GAGTACTGAGGGGTCAAATAACACTGAAGGAAGTGACACAATCACCTCCCATGCAGAATAAAACAATTTAT
7430 7440 7450 7460 7470 7480 7490

1310 1320 1330 1340 1350 1360 1370
AAACATGTGTCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATCAGTCGGACAAATTAGATGTTTCATCAA
|||||
AAACATGTGTCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATCAGTCGGACAAATTAGATGTTTCATCAA
7500 7510 7520 7530 7540 7550 7560

1380 1390 1400 1410 1420 1430 1440 1450
TATTACAGGGCTGCTATTAAACAAGAGATGGTGGTAATAACAACAATGGGTCCGAGATCTTCAGACCTGGAGG
|||||
TATTACAGGGCTGCTATTAAACAAGAGATGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACTTGGAGG
7570 7580 7590 7600 7610 7620 7630

1460 1470 1480 1490 1500 1510 1520
AGGAGATATGAGGGACAATTGGAGGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGT
|||||
AGGAGATATGAGGGACAATTGGAGGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGT
7640 7650 7660 7670 7680 7690 7700 7710

1530 1540 1550 1560 1570 1580 1590
AGCACCACCAAGGCAAGAGAGAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCT
|||||
AGCACCACCAAGGCAAGAGAGAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCT
7720 7730 7740 7750 7760 7770 7780

1600 1610 1620 1630 1640 1650 1660
TGGGTTCTTGGGAGCAGCAGGAACTATGGGCGCAGCCTCAATGACGCTGACGGTACAGGCCAGACAATT
|||||
TGGGTTCTTGGGAGCAGCAGGAACTATGGGCGCAGCCTCAATGACGCTGACGGTACAGGCCAGACAATT
7790 7800 7810 7820 7830 7840 7850

1670 1680 1690 1700 1710 1720 1730
ATTGTTCTGATATAGTGCAGCAGCAGAACTTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACT
|||||
ATTGTTCTGATATAGTGCAGCAGCAGAACTTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACT
7860 7870 7880 7890 7900 7910 7920

9. KUNZ-158-CL32L SEQ

HIVSC Human immunodeficiency virus type 1, isolate SC (3

LOCUS HIVSC 4273 bp ss-RNA VRL 15-JUN-1989
 DEFINITION Human immunodeficiency virus type 1, isolate SC (3' end of genome).
 ACCESSION M17450
 KEYWORDS
 SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate SC, proviral DNA.
 ORGANISM Human immunodeficiency virus type 1
 Viridae/ ss-RNA enveloped viruses; Retroviridae;
 Lentivirinae.
 REFERENCE 1 (bases 1 to 4273)
 AUTHORS Gungo, C., Guo, H.-B., Franchini, G., Aldovini, A., Collalti, E., Farrell, K., Wong-Steal, F., Gallo, R. C. and Reitz, M. S. Jr.
 TITLE Envelope sequences of two new United States HIV-1 isolates
 JOURNAL Virology 164, 531-536 (1988)
 STANDARD full staff review
 COMMENT Kindly made available in computer readable form by Marv Reitz, N. C. I., Bethesda, MD 20892 U.S.A. This isolate was taken from a California AIDS patient in 1984. There is an in-frame stop codon at position 3212 of the envelope coding sequence; the nef cds is uncertain beyond position 4049. A stop codon, 'taa,' in-frame with the nef sequence does exist at positions 4224-4226.

FEATURES	from	to/span	description
pept	1	330	vif protein (partial; AA at 1)
pept	270	560	vpr protein
pept	541	755	tat protein, exon 2 (first expressed exon)
	3088	3179	tat protein, exon 3 (AA at 3090)
pept	680	755	rev protein, exon 2 (first expressed exon)
	3088	3303	rev protein, exon 3 (AA at 3091)
pept	772	864	vpu protein (premature termination)
pept, ps	935	3505	envelope polyprotein (premature stop at 3212)
pept	3507	4226	nef protein
pre-mssg	1	4273	genomic mRNA
pre-mssg	1	4273	tat, rev, nef subgenomic mRNA
IVS	1	487	tat, rev, nef subgenomic mRNA intron 1
IVS	756	3088	tat cds intron 2
IVS	756	3088	rev cds intron 2
IVS	758	3088	tat, rev, nef subgenomic mRNA intron 2
LTR	3756	4273	3' LTR
rept	4248	4273	R repeat 3' copy
site	3212	3214	premature stop (tag) in env cds
BASE COUNT	1447 a	760 c	1053 g 1013 t

ORIGIN

Initial Score = 1158 Optimized Score = 2151 Significance = 0.00
 Residue Identity = 88% Matches = 2201 Mismatches = 214
 Gaps = 82 Conservative Substitutions = 0

```

X          10          20          30          40          50          60
AAGATCAG-AAGA-CAGTGGCAATGAGAGTGAAGGAGAA-ATATC----AGCACTTGTGGAGATGGGGGTGG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AATAGAGAAAGAGCAAGACAGTGGCAATGAGAGTGAAGGGATCAGGAAGGAATTAT-CAG-CACTTGTGG
810          920          930          940          950          960          970

          70          80          90          100          110          120          130
AAATGGGGCACCATGCTCCTTGGGATATTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTCTAT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGATGGGGCACCATGCTCCTTGGGATATTGATGATCTGTAGTGCTGCAGAACAAATTGTGGGTCACAGTCTAT
980          990          1000          1010          1020          1030          1040

          140          150          160          170          180          190          200
TATGGGGTGGGATGGGAAAGGAGCAACCACTCTATTTGTGCATCAGATGCTAAAGCATATGATACA

```

TATCGGGTACCTGTGTGGGAAGAGCAACCACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACA
 1050 1060 1070 1080 1090 1100 1110 1120
 210 220 230 240 250 260 270 280
 GAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTA
 GAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGGA
 1130 1140 1150 1160 1170 1180 1190
 290 300 310 320 330 340 350
 AATGTGACAGAAATTTTAACATGTGGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTA
 AATGTGACAGAAATTTTAACATGTGGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTA
 1200 1210 1220 1230 1240 1250 1260
 360 370 380 390 400 410 420
 TGGGATCAAGGCTAAGGCCATGTGTAAATTAACCCCACTCTGTGTTAGTTTAAAGTGCACCTGATTTGGGG
 TGGGATCAAGGCTAAGGCCATGTGTAAATTAACCCCACTCTGTGTTACTTTAAATTGCACTAATTTGAGG
 1270 1280 1290 1300 1310 1320 1330
 430 440 450 460 470 480 490
 AATGCTACTAATACCAATACTAGTAATACCAATAGTAGT-AGCGGGGAAATGATGATGGAGAAAGGAGAGAT
 AATGCTACTAATACCAATACTAGTAATACCAATAGTAGT-AGCGGGGAAATGATGATGGAGAAAGGAGAGAT
 1340 1350 1360 1370 1380 1390 1400
 500 510 520 530 540 550 560
 AAAAACTGCTCTTTCAATATCAGCACAAGNATAAGAGGTAAGGTGCAGAAAGAATATGCATTTTTTTTATAA
 GACAACTGCTCTTTCAATATCAGCACAAGCATAAGAAGTAAGGTACAGAAAGAATATGCATTTTTTTTATAA
 1410 1420 1430 1440 1450 1460 1470
 570 580 590 600 610 620 630 640
 ACTTGATATAATACCAATAGATAATGATACTACCAGCTATACGTTGACAAGTTGTAACACCTCAGTCATTAC
 ACTCGATGTAGTACCAATAGATAA-----TACGAGCTATACGTTGATAAATTGTAACACCTCAGTCATCAC
 1480 1490 1500 1510 1520 1530 1540
 650 660 670 680 690 700 710
 ACAGACCTATCCAAAGGTATCCTTTGAGCCAAATCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCT
 ACAGACCTATCCAAAGGTATCCTTTGAGCCAAATCCCATACACTATTGTGCCC--GGTGG-TTTGCGATTCT
 1550 1560 1570 1580 1590 1600 1610
 720 730 740 750 760 770 780
 AAAATGTAAATTAAGAACGTTCAATGGAAACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATGG
 AAAATGTAAATTAAGAACGTTCAATGGAAACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATGG
 1620 1630 1640 1650 1660 1670 1680
 790 800 810 820 830 840 850
 AATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCAGTCTAGCAGAAGAAGAGGTTAGTAATTAGATC
 AATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCAGTCTAGCAGAAGAAGAGGTTAGTACTTAGGTC
 1690 1700 1710 1720 1730 1740 1750
 860 870 880 890 900 910 920
 TGGCAATTTCAGAGACATGCTAAGACCATAAAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAG
 TGGCAATTTCAGAGACATGCTAAGACCATAAAGTACAGCTGAAGGAAGCTGTAGAAATTAATTGTACAAG
 1760 1770 1780 1790 1800 1810 1820
 930 940 950 960 970 980 990
 AATCAAGACAGATTCAGAGAGAGAGATTCAGTATCCAGAGAGGACCAAGGAGAGCATTGTTACAATAGGA--

[illegible]

```

|||||
AACAAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTC
2600      2610      2620      2630      2640      2650      2660

1770      1780      1790      1800      1810      1820      1830
CAGGCAAGGATATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTGCTCTGGA
|||||
CAGGCAAGGATGCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTGCTCTGGA
2670      2680      2690      2700      2710      2720      2730

1840      1850      1860      1870      1880      1890      1900
AAACTCATTGTGTCCTACTGCTGTGCTTGGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAAT
|||||
AAACTCATTTGACCACTACTGTGCTTGGGAATGCTAGTTGGAGTAATAAATCTTTGGACAAGATTTGGGGT
2740      2750      2760      2770      2780      2790      2800

1910      1920      1930      1940      1950      1960      1970
AACAATGACCTGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACATTCCTTAATTGAAGAA
|||||
AACAATGACCTGATGGAGTGGGAAAGAGAAATTAACAATTACACAAGCTTAATATACACCTTAATTGAAGAA
2810      2820      2830      2840      2850      2860      2870

1980      1990      2000      2010      2020      2030      2040      2050
TCGCAAAATCAGCAAGGAAAGAAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGG
|||||
TCGCAAGAACCAACCAAGGAAAGAAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGG
2880      2890      2900      2910      2920      2930      2940      2950

2060      2070      2080      2090      2100      2110      2120
TTTAACATAGCAATTTGGCTGTGGTATATAAAAAATATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGA
|||||
TTTAACATAGCAATTTGGCTGTGGTATATAAAAAATATTCATAATGATAGTAGGAGGCTTAGTAGGTTTAAGA
2960      2970      2980      2990      3000      3010      3020

2130      2140      2150      2160      2170      2180      2190
ATAGTTTTTGGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCAC
|||||
ATAGTTTTTGGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCGC
3030      3040      3050      3060      3070      3080      3090

2200      2210      2220      2230      2240      2250      2260
CTCCCAAGCCCAAGGAGACCCGACAGGCCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGA
|||||
CTCCCAAGCCCAAGGAGACCCGACAGGCCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGA
3100      3110      3120      3130      3140      3150      3160

2270      2280      2290      2300      2310      2320      2330
TCCATTGCAATTATGTAACGATCCTTAGCACTTATCTGGGACGATCTGCGGAGCCTTGTGCCTCTTCAGCTA
|||||
TCCGATGCAATTATGTAACGATCCTTAGCAATATCTGGGTCGACTAGCGGAGCC-TGTGCCTCTTCAGCTA
3170      3180      3190      3200      3210      3220      3230

2340      2350      2360      2370      2380      2390      2400      2410
CCACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACCTCTGGGACGCAGGGGGTGGGAAGC
|||||
CCACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACCTCTGGGACGCAGGGGGTGGGAAGC
3240      3250      3260      3270      3280      3290      3300

2420      2430      2440      2450      2460
CCTCAAAATATTGTGGAATCTCCTACAGTATTGGAGTCAGGAACTAAAG
|||||
CCTCAAAATATTGTGGAATCTCCTACAGTATTGGAGTCAGGAACTAAGG
3310      3320      3330      3340      3350      X

```


HIVMAL

Human immunodeficiency virus type 1, isolate MAL,

LOCUS HIVMAL 8229 bp ss-RNA VRL 15-JUN-1989
 DEFINITION Human immunodeficiency virus type 1, isolate MAL, complete genome.
 ACCESSION K03455
 KEYWORDS
 SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate MAL, proviral DNA, clone M-411.
 ORGANISM Human immunodeficiency virus type 1
 Viridae; ss-RNA enveloped viruses; Retroviridae;
 Lentivirinae.
 REFERENCE 1 (bases 1 to 8229)
 AUTHORS Alizon,M., Vain-Hobson,S., Montagnier,L. and Sonigo,P.
 TITLE Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isolates from African patients
 JOURNAL Cell 46, 63-74 (1986)
 STANDARD full staff_review
 COMMENT A region similar to the vpu cds of HIVELI exists from positions 5636 (starts with 'eta' instead of 'atg') to 5881. To date it has not been possible to construct an infectious clone for the MAL viral genome (Keith Peden, Johns Hopkins University School of Medicine, Baltimore, MD).
 FEATURES

	from	to/span	description
pept	350	1667	gag polyprotein
pept	(1663	4671	pol polyprotein (NH2-terminus uncertain; AA at 1663)
pept	4616	5184	vif protein
pept	5134	5424	vpr protein
pept	5405	5519	tat protein, exon 2 (first expressed exon)
	7556	8007	tat protein, exon 3 (AA at 7960)
pept	5544	5619	rev protein, exon 2 (first expressed exon)
	7958	8236	rev protein, exon 3 (AA at 7961)
pept	(5636	5881	vpu protein (AA at 5636)
pept	5799	6378	envelope polyprotein precursor
pept	6380	8008	nef protein
pre-msg	1	8228	genomic mRNA
pre-msg	1	8228	tat, rev, nef subgenomic mRNA
IVS	308	5351	tat, rev, nef subgenomic mRNA intron 1
IVS	5620	7856	tat cds intron 2
IVS	5620	7858	rev cds intron 2
IVS	5620	7858	tat, rev, nef subgenomic mRNA intron 2
LTR	(1	177	5' LTR
LTR	8576) 8229	3' LTR
rep	(1	35	R repeat 5' copy
rep	8133	8229	R repeat 3' copy
binding	175	186	primer (Lys-tRNA) binding site
signal	8204	8208	mRNA polyadenylation signal
BASE COUNT	3335 a	1627 c	2204 g 2043 t
ORIGIN	Cap site of genomic RNA.		

Initial Score = 916 Optimized Score = 2060 Significance = 0.00
 Residue Identity = 84% Matches = 2085 Mismatches = 350
 Gaps = 43 Conservative Substitutions = 0

```

X      10      20      30      40      50      60      70
AGAGACAGAGACAGTGGCAATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAATGGG
|||||
AGAGACAGAGACAGTGGCAATGAGAGTGAAGGAGATACA-GAGGAATTATCAAAA--CTGGTGGAGATGGG
5780      5790      5800      5810      5820      5830      5840

      80      90      100      110      120      130      140
GCACCATGCTCTTGGGATATTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTCTATTATGGGG
|||||
GCATGATGCTCTTGGGATATTGATGACCTGTAGTATTGCAGAGATTGTGGGTTACAGTTTATTATGGGG
5850      5860      5870      5880      5890      5900      5910

```

150 160 170 180 190 200 210
 TACCTGTGTCGAGAGGAGCAACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTAC
 TACCTGTGTCGAGAGGAGCAACCACTCTATTTTGTGCATCAGATGCTAAATCATATGAAACAGAAGTAC
 5920 5930 5940 5950 5960 5970 5980 5990

220 230 240 250 260 270 280
 ATAAATGTTTGGGCAACACATGCTGTGTACCCACAGACCCCAACCCACAAGAGTAGTATTGGTAAATGTGA
 ATAAATGTTTGGGCAACACATGCTGTGTACCCACAGACCCCAACCCACAAGAAATAGAACTGGAAATGTCA
 6000 6010 6020 6030 6040 6050 6060

290 300 310 320 330 340 350 360
 CAGAAATTTTAAACATGTGGGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATC
 CAGAAAGGTTTAAACATGTGGGAAAAATAACATGGTGGAGCAGATGCATGAGGATATAATCAGTTTATGGGATC
 6070 6080 6090 6100 6110 6120 6130

370 380 390 400 410 420
 AAAGCCTAAAGCAGATGTGTAAATTAACCCCACTCTGTGTTAGTTTAAAGTGCCTGATTTG---GGGAATG
 AAAGCCTAAAGCAGATGTGTAAATTAACCCCACTCTGTGTTAGTTTAAAGTGCCTGATTTG---GGGAATG
 6140 6150 6160 6170 6180 6190 6200

430 440 450 460 470 480 490
 CTACTAAT---ACCAATACTAATAATACCAATAGTAGTAGCGGGGAATGATGATGGAGAAAGGAGAGATAA
 CTACTAAT---ACCAATACTAATAATACCAATAGTAGTAGCGGGGAATGATGATGGAGAAAGGAGAGATAA
 6210 6220 6230 6240 6250 6260 6270

500 510 520 530 540 550 560 570
 AAAACTGCTCTTTCAATATCAGCACAAGNATAAGAGGTAAAGGTGCAGAAAGAATATGCATTTTTTTATAAAC
 AAAACTGCTCTTTCAATATCAGCACAAGNATAAGAGGTAAAGGTGCAGAAAGAATATGCATTTTTTTATAAAC
 6280 6290 6300 6310 6320 6330 6340

580 590 600 610 620 630
 TTGATATAATACCAATA---GATAATGATACTACCAGCTATACGTTGACAAGTTGTAACACCTCAGTCATTA
 TTGATATAATACCAATA---GATAATGATACTACCAGCTATACGTTGACAAGTTGTAACACCTCAGTCATTA
 6350 6360 6370 6380 6390 6400 6410

640 650 660 670 680 690 700 710
 CACAGGCTGTCTCAAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCGGCTGGTTTTGCGATTC
 CACAGGCTGTCTCAAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCGGCTGGTTTTGCGATTC
 6420 6430 6440 6450 6460 6470 6480

720 730 740 750 760 770 780
 TAAATGTATAATAAGACGTTCAATGGAACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATG
 TAAATGTATAATAAGACGTTCAATGGAACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATG
 6490 6500 6510 6520 6530 6540 6550 6560

790 800 810 820 830 840 850
 GAATTAGGCCAGTATGTAACCACTCAACTGCTGTTGAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGAT
 GAATTAGGCCAGTATGTAACCACTCAACTGCTGTTGAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGAT
 6570 6580 6590 6600 6610 6620 6630

860 870 880 890 900 910 920
 CTGCAATTTTCAAGCAATGCTTAAACCATATAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAA
 CTGCAATTTTCAAGCAATGCTTAAACCATATAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAA
 6640 6650 6660 6670 6680 6690 6700

800 810 820 830 840 850 860

GACCCGACACACATACAGAAAAAETATCGTATCCAGAGGGGACCAGGGAGGCATTGTGTTACAATAGGAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GGCCTGGAAACATACAGAGAGAGGATACATTTTC-----GGCCCAGGGCAAGCACTCTATACAACAGGGA
6710 6720 6730 6740 6750 6760 6770

1000 1010 1020 1030 1040 1050 1060 1070

(AATAGGAAAT)ATGAGACAGCACATTGTACATTAGTAGAGCAAATGCAATGCCACTTTAAAACAGATAG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TASTGGGAGATAAAGAAGAACATATTGTACTATTAAATGAAACAGAATGGGATAAACTTTACAACAGGTAG
6780 6790 6800 6810 6820 6830 6840

1080 1090 1100 1110 1120 1130 1140

CTAGCAAACTAAGAGAACATTTGGAAATAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CTGTAAACTAGGA-AGC--CTTCTTAACAAAACAAAATAATTTTTTAATTCATCCTCAGGAGGGGACCCAG
6850 6860 6870 6880 6890 6900 6910

1150 1160 1170 1180 1190 1200 1210

(AATTGTANCGCACAGTTTTAATTGTGGAGGGGAATTTTTCTACTGTAATTCAACACAACTGTTTAATAGTA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AAATTACANCACACAGTTTTTAATTGTAGAGGGGAATTTTTCTACTGTAATACATCAAACTGTTTAATAGTA
6920 6930 6940 6950 6960 6970 6980

1220 1230 1240 1250 1260 1270 1280

CTTGTTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACACTGAAGGAAGTGACACAATCACACTCCCAT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CATGGUASNAATAT-GGTGCAGCACT-AAG----TAATAGCACAGAGTCAACTGGTAGTATCACACTCCCAT
6990 7000 7010 7020 7030 7040 7050

1290 1300 1310 1320 1330 1340 1350

GCAGNATAAAACAATTTATAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATCAGCGGAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCAGNATAAAACAATTTATAAATATGTGGCAGAAAACAGGAAAAGCTATGTATGCCCTCCCATCGCAGGAG
7060 7070 7080 7090 7100 7110 7120

1360 1370 1380 1390 1400 1410 1420

(AATTAGATGTTTCATCAATATTACAGGGCTGCTATTACAAGAGATGGTGGTAATAACA---ACAAT-GGG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TCATCAACTGTTTATCAATATTACAGGGCTGATATTAACAAGAGATGGTGGAAATAGTAGTGACAATAGTG
7130 7140 7150 7160 7170 7180 7190

1430 1440 1450 1460 1470 1480 1490

TC---CGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACAATGAGACCTTACAGACCTGGAGGAGGAGATATGAGGGACAATTGGATAAGTGAATTATATAAATATAAAG
7200 7210 7220 7230 7240 7250 7260

1500 1510 1520 1530 1540 1550 1560

TAGTAAAAATTGGACCATTAGAGTAGCACCACCAAGGCAAGAGAGAGTGGTGACAGAGAGAAAAAGAG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TAGTAAAGATTGACCCCTAGAGTAGCACCACCAAGGCAAGAGAGAGTGGTGGAAAGAGAAAAAGAG
7270 7280 7290 7300 7310 7320 7330

1570 1580 1590 1600 1610 1620 1630 1640

CAGTGGGAYTAGAGCTTTGTTCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CAATAGGACTAGAGCCATGTTCTTGGGTTCTTGGGAGCAGCAGGAAGCAGATGGGCGCAGCGTCACTAA
7340 7350 7360 7370 7380 7390 7400 7410

1650 1660 1670 1680 1690 1700 1710

CCCTAGCGTACAGGCCAGACAATTATTGTCTGGTATAGTGACAGCAGCAGAACAAATTTGCTGAGGGCTATTG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CCCTAGCGTACAGGCCAGACATTACTGTCTGGTATAGTGCAACAGCAAAACAATTTGCTGAGGGCTATAG
7420 7430 7440 7450 7460 7470 7480

1720 1730 1740 1750 1760 1770 1780
AGGCGCAACACACATCTCTTGAACCTCACAGTCTGGGGCATCAAGCAGCTCCAGGGCAAGAATCCTGGCTGTGG
|||||
AGGCGCAACACACATCTCTTGAACCTCACAGTCTGGGGCATTAACAGCTCCAGGGCAAGAGTCTGGCTGTGG
7490 7500 7510 7520 7530 7540 7550

1760 1800 1810 1820 1830 1840 1850
AAGGATACCTAAGGATCAACAGCTCTGGGGATTGGGGTTGCTCTGGAAACTCATTTCACCACTGCTG
|||||
AAGGATACCTAAGGATCAACAGCTCTAAGGATGGGGTTGCTCTGGAAACACATTTCACCACTGCTG
7560 7570 7580 7590 7600 7610 7620

1860 1870 1880 1890 1900 1910 1920
TGCCTTGGATCTAGTTGGAGTAATAATCTCTGGAACAGATTTGGAATAACATGACCTGGATGGAGTGGG
|||||
TGCCTTGGATCTAGTTGGAGTAATAGATCTCTAGATGACATTTGGAATAATATGACCTGGATGCAGTGGG
7630 7640 7650 7660 7670 7680 7690

1930 1940 1950 1960 1970 1980 1990 2000
ACAGGGAATTTACCAATTACACAGCTTAATACATTCCTTAATTGAAGAATCGCAAACCCAGCAAGAAAAGA
|||||
AAGGAGAACTTAGCAATTACACAGGCTAATATACAACCTTAATTGAAGAATCGCAAATCCAGCAAGAAAAGA
7700 7710 7720 7730 7740 7750 7760 7770

2010 2020 2030 2040 2050 2060 2070
ATGACCAAGGATTTATTTGGAAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGT
|||||
ATGACCAAGGATTTATTTGGAAATTGGACAGTGGGCAAGTTTGTGGAATTGGTTTAGCATATCAAATGGCTGT
7780 7790 7800 7810 7820 7830 7840

2080 2090 2100 2110 2120 2130 2140
GGATATAGCAATTTATTAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGTCTGTACTTTCTA
|||||
GGATATAGCAATTTATTAATGATAGTAGGAGGCTTAATAGGTTTAAGAATAATTTTTGTCTGTGCTTTCTT
7850 7860 7870 7880 7890 7900 7910

2150 2160 2170 2180 2190 2200 2210
TACGGAATAGGATAGGCAAGGATTTACCATTTATCTTTTACAGACCCACCTCCCAACCCCGAGGGG---AC
|||||
TACGGAATAGGATAGGCAAGGATTTACCATTTATCTTTTACAGACCCCTCCTCCCAACCCCGAGGGGACCAC
7920 7930 7940 7950 7960 7970 7980

2220 2230 2240 2250 2260 2270 2280
CCGACAGGCCCCAAGGATAGAGAGAGGAGGTGGAGAGAGAGACAGAGACAGATCCATTGATTAGTGAACG
|||||
CCGACAGGCCCCAAGGATAGAGAGAGGAGGTGGAGAGCAAGGACAGAGGACAGATCAATTGATTGGTGAACG
7990 8000 8010 8020 8030 8040 8050

2290 2300 2310 2320 2330 2340 2350
GATCTTAGCACTTATCTGGGACGATCTGGGAGGCTTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTA
|||||
GATCTTAGCACTTATCTGGGACGACCTGAGGAACC-TGTGCCTCTTCAGTTACCACCGCTTGAGAGACTTA
8060 8070 8080 8090 8100 8110 8120

2360 2370 2380 2390 2400 2410 2420 2430
CTCTTGATTTATAACGAGGATTTGGAACCTCTGGGACGCAGGGGGTGGGAAGCCCTCAAATATTGGTGGAA
|||||
CTCTTAATTGCAACGAGGATTTGGAACCTCTGGGACGCAGGGGGTGGGAAGCCCTCAAATATCTGTGGAA
8130 8140 8150 8160 8170 8180 8190 8200

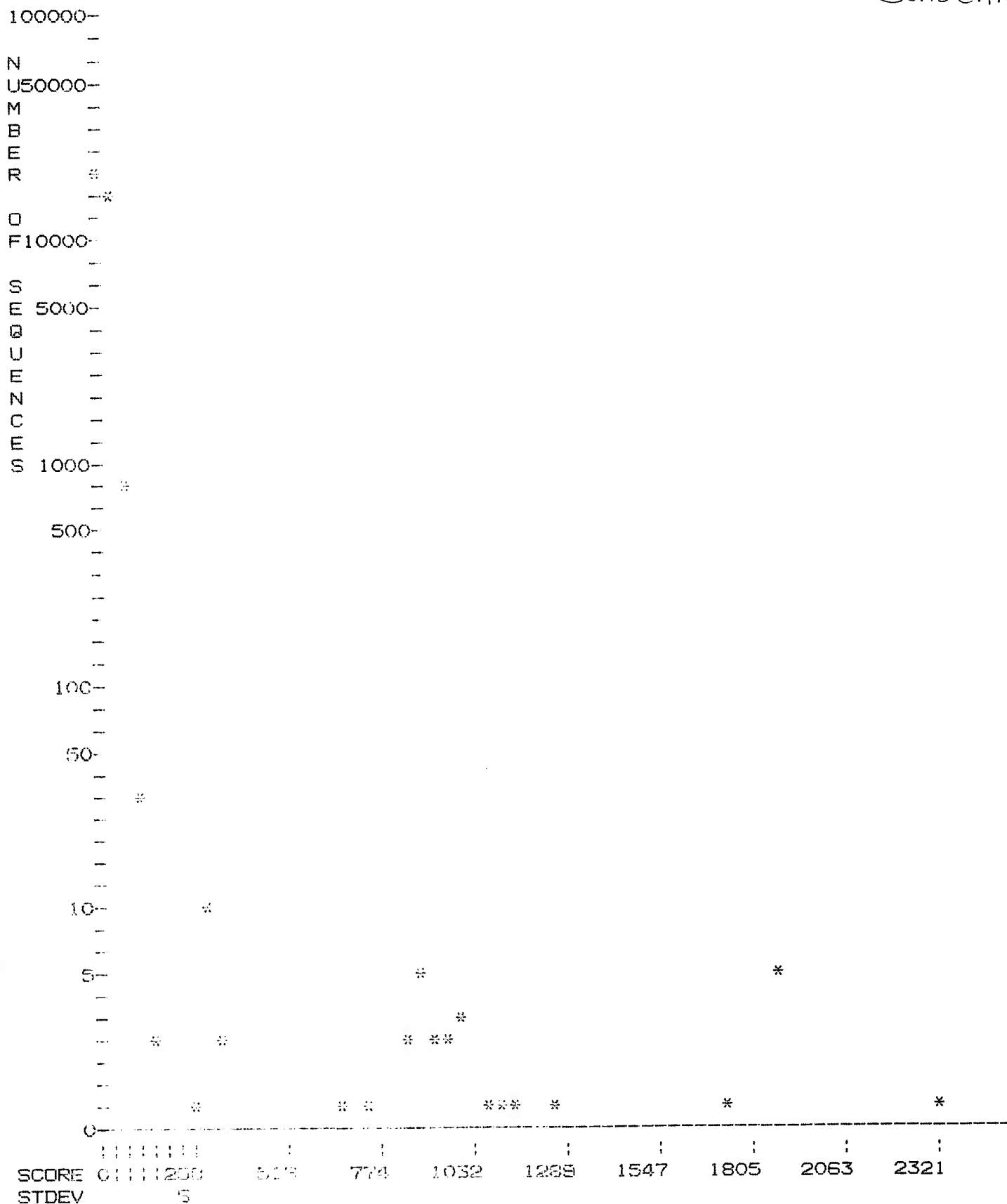
2440 2450 X
CTCTTACATTTATTTGGGATCAAGGACTAAG
|||||
CTCTTACATTTATTTGGGATCAAGGACTAAG
8210 8220 8230

Query sequence being compared: KUNZ-158-CL33.SEQ
 Number of sequences searched: 31228
 Number of scores above cutoff: 38

Kunz
 07/158652

Results of the initial comparison of KUNZ-158-CL33.SEQ with: Claim 33
 Data bank: GenBank 62.0: all entries

Genbank



PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	146		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	37	36	19.78

Times:	CPU	Total Elapsed
	00:53:20.05	02:16:33.00

Number of peptides:	37183850
Number of sequences searched:	31228
Number of scores above cutoff:	38

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 115 standard deviations above mean ****					
1. HIVBRUCB	Human immunodeficiency virus t	9229	2321	2436	115.50	0
	**** 93 standard deviations above mean ****					
2. HIVPV22	Human immunodeficiency virus t	9770	1877	2180	93.04	0
	**** 92 standard deviations above mean ****					
3. HIVKXBY	Human immunodeficiency virus t	3156	1873	2176	92.84	0
4. HIVGH102	Human immunodeficiency virus t	8932	1872	2176	92.79	0
5. HIVKX22C3	Human immunodeficiency virus t	9718	1858	2164	92.08	0
	**** 85 standard deviations above mean ****					
6. HIVNL43	Human immunodeficiency virus t	9709	1729	2169	85.56	0
	**** 61 standard deviations above mean ****					
7. HIVELICG	Human immunodeficiency virus t	9176	1246	1893	61.14	0
	**** 56 standard deviations above mean ****					
8. HIVSC	Human immunodeficiency virus t	4273	1159	2139	56.74	0
	**** 54 standard deviations above mean ****					
9. HIVH3PHE	Human immunodeficiency virus t	3563	1112	1761	54.36	0
	**** 52 standard deviations above mean ****					
10. HIVZ301	Human immunodeficiency virus t	3457	1066	1959	52.03	0
	**** 46 standard deviations above mean ****					
11. HIVMNCB	Human immunodeficiency virus t	9738	996	2203	48.49	0
	**** 47 standard deviations above mean ****					
12. HIVZ22C	Human immunodeficiency virus t	9081	983	1908	47.84	0
13. HIVZ6	Human immunodeficiency virus t	5159	983	1915	47.84	0
14. HIVTH32	Human immunodeficiency virus t	2903	975	1365	47.43	0
15. HIVJY1	Human immunodeficiency virus t	2653	972	1917	47.28	0
	**** 45 standard deviations above mean ****					
16. HIVCDC42	Human immunodeficiency virus t	3373	932	2191	45.26	0
	**** 44 standard deviations above mean ****					
17. HIVMAL	Human immunodeficiency virus t	9229	916	2041	44.45	0
	**** 42 standard deviations above mean ****					

18.	HIVSFENV	Human immunodeficiency virus ty	2622	887	1567	42.98	0
19.	HIVRF	Human immunodeficiency virus t	9128	887	1567	42.98	0
20.	HIVSF205	Human immunodeficiency virus t	9737	883	1941	42.78	0

The scores below are sorted by optimized score.
Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. HIVBRUC8	Human immunodeficiency virus t	9229	2321	2436	0.00	0
2. HIVMNC9	Human immunodeficiency virus t	9738	996	2203	0.00	0
3. HIVCDC42	Human immunodeficiency virus t	3373	932	2191	0.00	0
4. HIVPV22	Human immunodeficiency virus t	9770	1877	2180	0.00	0
5. HIVBH102	Human immunodeficiency virus t	8932	1872	2176	0.00	0
6. HIVHXB3	Human immunodeficiency virus t	3156	1873	2176	0.00	0
7. HIVNL43	Human immunodeficiency virus t	9709	1729	2169	0.00	0
8. HIVHXBRC6	Human immunodeficiency virus t	9718	1858	2164	0.00	0
9. HIVS6	Human immunodeficiency virus t	4273	1159	2139	0.00	0
10. HIVMAL	Human immunodeficiency virus t	9229	916	2041	0.00	0
11. HIVBRV3	Human immunodeficiency virus t	3600	861	2027	0.00	0
12. HIVZ321	Human immunodeficiency virus t	3457	1066	1959	0.00	0
13. HIVSF208	Human immunodeficiency virus t	9737	883	1941	0.00	0
14. HIVJY1	Human immunodeficiency virus t	2653	972	1917	0.00	0
15. HIVZE	Human immunodeficiency virus t	5159	983	1915	0.00	0
16. HIVZ2Z8	Human immunodeficiency virus t	9081	983	1908	0.00	0
17. HIVEL109	Human immunodeficiency virus t	9176	1246	1893	0.00	0
18. HIVH3BH8	Human immunodeficiency virus t	3563	1112	1761	0.00	0
19. HIVSFENV	Human immunodeficiency virus ty	2622	887	1567	0.00	0
20. HIVRF	Human immunodeficiency virus t	9128	887	1567	0.00	0

1. KUNZ-158-CL33. SE2

HIVBRUC8 Human immunodeficiency virus type 1, isolate BRU,

LOCUS HIVBRUC8 9229 bp ss-RNA VRL 15-JUN-1989
DEFINITION Human immunodeficiency virus type 1, isolate BRU, complete genome (LAV-1).
ACCESSION K02013
KEYWORDS TAP protein; TAT protein; acquired immune deficiency syndrome; complete genome; env gene; gag gene; long terminal repeat; pol gene; polyprotein; provirus; reverse transcriptase; trans-activator.
SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate BRU (LAV-1), proviral DNA clone lambda-J19.
ORGANISM Human immunodeficiency virus type 1
Virus: ss-RNA enveloped viruses; Retroviridae; Lentivirinae.
REFERENCE 1 (bases 1 to 9229)
AUTHORS Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. and Alizon, M.
TITLE Nucleotide sequence of the AIDS Virus, LAV
JOURNAL Cell 40, 9-17 (1985)
STANDARD full_staff_review
REFERENCE 2 (bases 1712 to 1749; revision of [1])
AUTHORS Alizon, M., Wain-Hobson, S., Montagnier, L. and Sonigo, P.
TITLE Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isolates from African patients
JOURNAL Cell 46, 63-74 (1986)
STANDARD full_staff_review
COMMENT The original LAV, sometimes called LAV-1 to distinguish it from HIV2 (LAV-2), is now referred to as HIV-1bru. An infectious clone

of this virus has been constructed by Keith Peden, Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205 (301) 955-3652. HIVNL43 is also an infectious clone having for its 3' half a clone of the BRU isolate. Acquired immune deficiency syndrome (AIDS) is caused by a retrovirus known by several different names, probably representing two separate strains: human T-cell lymphotropic virus-III (HTLV-III) and lymphadenopathy-associated virus (LAV) are thought to be one strain, and AIDS-associated retrovirus type 2 (ARV-2) the other. All three viruses, whose sequences do not differ by more than about 6%, are believed to belong to the retroviral subfamily Lentiviridae, or "slow" viruses. For the details of the annotation and for other pertinent references, see the HIV reference entry.

FEATURES	from	to/span	description
pept	336	1874	gag polyprotein
pept	(1631	4578	pol polyprotein (NH2-terminus uncertain; AA at 1631)
pept	4623	5201	vif protein
pept	5141	5431	vpr protein
pept	5412	5626	tat protein, exon 2 (first expressed exon)
	7972	8017	tat protein, exon 3 (AA at 7973)
pept	5551	5626	rev protein, exon 2 (first expressed exon)
	7972	8246	rev protein, exon 3 (AA at 7974)
pept	5643	5838	vpu protein
pept	5803	8388	envelope polyprotein
pept	8380	9010	nef protein
pre-msg	1	9229	genomic mRNA
pre-msg	1	9229	tat, rev, nef subgenomic mRNA
IVS	290	5358	tat, rev, nef subgenomic mRNA intron 1
IVS	5627	7971	tat cds intron 2
IVS	5627	7971	rev cds intron 2
IVS	5627	7971	tat, rev, nef subgenomic mRNA intron 2
LTR	(1	180	5' LTR
LTR	8675) 9223	3' LTR
rpt	(1	97	R repeat 5' copy
rpt	8133	8223	R repeat 3' copy
binding	182	189	primer (Lys-tRNA) binding site
revision	1712	1743	atttcttcagagcagaccagagccaacagccccaccag in [2]; ag in [1]
signal	8205	8210	mRNA polyadenylation signal
BASE COUNT	3269 a	1656 c	2232 g 2052 t
ORIGIN	Cap site of genomic RNA.		

Initial Score = 2321 Optimized Score = 2436 Significance = 0.00
 Residue Identity = 99% Matches = 2436 Mismatches = 3
 Gaps = 1 Conservative Substitutions = 0

```

X      10      20      30      40      50      60      70
ATGAGAGTGAAGGAGAAATATCAGCAGCTTGTGGAGATGGGGGTGGAAATGGGGCACCATGCTCCTTGGGATA
|||||
ATGAGAGTGAAGGAGAAATATCAGCAGCTTGTGGAGATGGGGGTGGAAATGGGGCACCATGCTCCTTGGGATA
X  5510      5820      5830      5840      5850      5860      5870

      80      90      100      110      120      130      140
TTGATGATCTGTAGTCTACAGAAAAATTTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCA
|||||
TTEATGATCTGTAGTCTACAGAAAAATTTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCA
5880      5890      5900      5910      5920      5930      5940

      150      160      170      180      190      200      210
ACCAACACGCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACAT
|||||
ACCAACACGCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACAT
5950      5960      5970      5980      5990      6000      6010

```

220 230 240 250 260 270 280
 GCCTGTGTACCCACAGACCCCAACCCACAGAGTAGTATTGGTAAATGTGACAGAAAATTTTAACATGTGG
 |||||
 GCCTGTGTACCCACAGACCCCAACCCACAGAGTAGTATTGGTAAATGTGACAGAAAATTTTAACATGTGG
 6020 6030 6040 6050 6060 6070 6080 6090

290 300 310 320 330 340 350 360
 AAAATATGACATGCTATACAGATGATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTA
 |||||
 AAAATATGACATGCTATACAGATGATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTA
 6100 6110 6120 6130 6140 6150 6160

370 380 390 400 410 420 430
 AAATTAACUCCCACTCTGTGTAGTCTAAAGTGCACCTGATTTGGGGAATGCTACTAATACCAATACTAGTAAT
 |||||
 AAATTAACUCCCACTCTGTGTAGTCTAAAGTGCACCTGATTTGGGGAATGCTACTAATACCAATACTAGTAAT
 6170 6180 6190 6200 6210 6220 6230

440 450 460 470 480 490 500
 ACCAATAGTAGTAGCCGGGAAATGATGATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACA
 |||||
 ACCAATAGTAGTAGCCGGGAAATGATGATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACA
 6240 6250 6260 6270 6280 6290 6300

510 520 530 540 550 560 570
 AGNATAAGAGGTAAAGTGCAGAAAGAAATATGCATTTTTTTTATAAACTTGATATAATACCAATAGATAATGAT
 |||||
 AGNATAAGAGGTAAAGTGCAGAAAGAAATATGCATTTTTTTTATAAACTTGATATAATACCAATAGATAATGAT
 6310 6320 6330 6340 6350 6360 6370

580 590 600 610 620 630 640
 ACTACCAGCTATACGTTGACAAAGTTGTAAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAG
 |||||
 ACTACCAGCTATACGTTGACAAAGTTGTAAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAG
 6380 6390 6400 6410 6420 6430 6440 6450

650 660 670 680 690 700 710 720
 CCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTCGATTCTAAAATGTAATAATAAGACGTTCAATGGA
 |||||
 CCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTCGATTCTAAAATGTAATAATAAGACGTTCAATGGA
 6460 6470 6480 6490 6500 6510 6520

730 740 750 760 770 780 790
 ACAGGACCATGTACAAATGTACGACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTG
 |||||
 ACAGGACCATGTACAAATGTACGACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTG
 6530 6540 6550 6560 6570 6580 6590

800 810 820 830 840 850 860
 CTGTTGAATGBCAGTCTAGCAGAAAGAGAGGTAGTAATTAGATCTGCCAATTTTCACAGACAATGCTAAAACC
 |||||
 CTGTTGAATGBCAGTCTAGCAGAAAGAGAGGTAGTAATTAGATCTGCCAATTTTCACAGACAATGCTAAAACC
 6600 6610 6620 6630 6640 6650 6660

870 880 890 900 910 920 930
 ATAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGTATC
 |||||
 ATAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGTATC
 6670 6680 6690 6700 6710 6720 6730

940 950 960 970 980 990 1000
 CGTATCCAGAGGGGACGAGGAGAGCAATTTGTTACAATAGGAAAAATAGGAAATATGAGACAAGCACATTGT
 |||||
 CGTATCCAGAGGGGACGAGGAGAGCAATTTGTTACAATAGGAAAAATAGGAAATATGAGACAAGCACATTGT
 6740 6750 6760 6770 6780 6790 6800 6810

1010 1020 1030 1040 1050 1060 1070 1080
 AACATTAGTAGAGCAAAATGCAATGCCACTTTAAACAGATAGCTAGCAAATTAAGAGAACAATTTGGAAAT
 AACTTAGTAGAGCAAAATGCAATGCCACTTTAAACAGATAGCTAGCAAATTAAGAGAACAATTTGGAAAT
 6820 6830 6840 6850 6860 6870 6880
 1090 1100 1110 1120 1130 1140 1150
 AATAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTAAATTGTGGA
 AATAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTAAATTGTGGA
 6890 6900 6910 6920 6930 6940 6950
 1160 1170 1180 1190 1200 1210 1220
 GGGGAATTTTTCTACTGTAATTCACACAACTGTTTAATAGTACTTGTTTAAATAGTACTTGGAGTACTGAA
 GGGGAATTTTTCTACTGTAATTCACACAACTGTTTAATAGTACTTGTTTAAATAGTACTTGGAGTACTGAA
 6960 6970 6980 6990 7000 7010 7020
 1230 1240 1250 1260 1270 1280 1290
 GGGTCAAAATAGCACTGAGGAACTGACACAATCAGACTCCCATGCAGAATAAAACAATTTATAACATGTGG
 GGGTCAAAATAGCACTGAGGAACTGACACAATCAGACTCCCATGCAGAATAAAACAATTTATAACATGTGG
 7030 7040 7050 7060 7070 7080 7090
 1300 1310 1320 1330 1340 1350 1360
 CAGGAAGTAGGAAAAGCAATGTATGCCCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAAATATTACAGGG
 CAGGAAGTAGGAAAAGCAATGTATGCCCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAAATATTACAGGG
 7100 7110 7120 7130 7140 7150 7160 7170
 1370 1380 1390 1400 1410 1420 1430 1440
 CTGCTATTACAGAGAGATGGTGGTAATAACAACAATGGGTCCGAGATCTTCAGACCTGGAGGAGGAGATATG
 CTGCTATTACAGAGAGATGGTGGTAATAACAACAATGGGTCCGAGATCTTCAGACCTGGAGGAGGAGATATG
 7180 7190 7200 7210 7220 7230 7240
 1450 1460 1470 1480 1490 1500 1510
 AGGGACAATTGAGAGAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACC
 AGGGACAATTGAGAGAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACC
 7250 7260 7270 7280 7290 7300 7310
 1520 1530 1540 1550 1560 1570 1580
 AAGGCAAAAGAGAGAGTGGTGCAAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCCTTGGGTTCTTG
 AAGGCAAAAGAGAGAGTGGTGCAAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCCTTGGGTTCTTG
 7320 7330 7340 7350 7360 7370 7380
 1590 1600 1610 1620 1630 1640 1650
 GGAGCAGCAGGAAAGCACTATGGGCGCACGGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGT
 GGAGCAGCAGGAAAGCACTATGGGCGCACGGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGT
 7390 7400 7410 7420 7430 7440 7450
 1660 1670 1680 1690 1700 1710 1720
 ATAGTGCAGCAGCAGAAACAATTTGCTGAGGGCTATTGAGGGCGCAACAGCATCTGTTGCAACTCACAGTCTGG
 ATAGTGCAGCAGCAGAAACAATTTGCTGAGGGCTATTGAGGGCGCAACAGCATCTGTTGCAACTCACAGTCTGG
 7460 7470 7480 7490 7500 7510 7520 7530
 1730 1740 1750 1760 1770 1780 1790 1800
 GGCATCAAGCACTCCAGGGCAAGAAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATT
 GGCATCAAGCACTCCAGGGCAAGAAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATT
 7540 7550 7560 7570 7580 7590 7600

1810 1820 1830 1840 1850 1860 1870
 TGGGATTGCTCTGGAAGAACTCATTTCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTG
 TGGGATTGCTCTGGAAGAACTCATTTCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTG
 7610 7620 7630 7640 7650 7660 7670
 1880 1890 1900 1910 1920 1930 1940
 GAACAGATTTGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACAT
 GAACAGATTTGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACAT
 7680 7690 7700 7710 7720 7730 7740
 1950 1960 1970 1980 1990 2000 2010
 TCCTTAATTGAAGAATTCGCAAAACCAAGCAAGAAAGAAATGAACAAGAATTATTGGAATTAGATAAATGGGCA
 TCCTTAATTGAAGAATTCGCAAAACCAAGCAAGAAAGAAATGAACAAGAATTATTGGAATTAGATAAATGGGCA
 7750 7760 7770 7780 7790 7800 7810
 2020 2030 2040 2050 2060 2070 2080
 AGTTTGTGGAAAT TGGTTTAACATAACAAATTGGCTGTGGTATATAAAAATATTCATAATGATAGTAGGAGGC
 AGTTTGTGGAAAT TGGTTTAACATAACAAATTGGCTGTGGTATATAAAAATATTCATAATGATAGTAGGAGGC
 7820 7830 7840 7850 7860 7870 7880 7890
 2090 2100 2110 2120 2130 2140 2150 2160
 TTGGTAGGCTTAAGAATAGT TTTTCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTA
 TTGGTAGGCTTAAGAATAGT TTTTCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTA
 7900 7910 7920 7930 7940 7950 7960
 2170 2180 2190 2200 2210 2220 2230
 TCGTTTCAGACCCACCTCCCAACCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAG
 TCGTTTCAGACCCACCTCCCAACCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAG
 7970 7980 7990 8000 8010 8020 8030
 2240 2250 2260 2270 2280 2290 2300
 AGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGCCTT
 AGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGCCTT
 8040 8050 8060 8070 8080 8090 8100
 2310 2320 2330 2340 2350 2360 2370
 GTGCCTCTTCAGCTACCAACGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACCTTCTGGGACG
 GTGCCTCTTCAGCTACCAACGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACCTTCTGGGACG
 8110 8120 8130 8140 8150 8160 8170
 2380 2390 2400 2410 2420 2430 X
 CAGGGGGTGGGAAGCCCTCAAATATTGGTGGAACTCTCTACAGTATTGGAGTCAGGAACTAAAG
 CAGGGGGTGGGAAGCCCTCAAATATTGGTGGAACTCTCTACAGTATTGGAGTCAGGAACTAAAG
 8180 8190 8200 8210 8220 8230 8240

2. KUNZ-158-CL33. SEQ

HIVMNCE Human immunodeficiency virus type 1, isolate MN, c

LOCUS	HIVMNCE	8738 bp ss-RNA	VRL	15-JUN-1989
DEFINITION	Human immunodeficiency virus type 1, isolate MN, complete genome.			
ACCESSION	M17449			
KEYWORDS	.			
SOURCE	Human immunodeficiency virus type 1 (HIV-1), isolate MN, proviral DNA.			
ORGANISM	Human immunodeficiency virus type 1 Virus; ss-RNA enveloped viruses; Retroviridae;			

REFERENCE 1 (bases 1 to 9738)
 AUTHORS Gorgo, C., Guo, H.-G., Franchini, G., Aldovini, A., Collalti, E.,
 Farrell, K., Wong-Staal, F., Gallo, R. C. and Reitz, M. S. Jr.
 TITLE Envelope sequences of two new United States HIV-1 isolates
 JOURNAL Virology 164, 531-536 (1988)
 STANDARD full staff_review
 COMMENT Sequence kindly provided in computer readable form by Marv Reitz,
 R.C.I., Bethesda, MD. 20892 U. S. A.

The MN isolate was taken from a pediatric AIDS patient in 1984.
 The pol coding sequence shows an in-frame stop codon at position
 3783; the nef protein is prematurely truncated at position 9357.
 The vpu protein is prematurely truncated at position 6142.

FEATURES	from	to/span	description
pept	787	2307	gag polyprotein
pept. ps (2031	5111	pol polyprotein (NH2-terminus uncertain; AA at 2031; in-frame stop codon at 3783)
pept	5056	5634	vif protein
pept	5574	5864	vpr protein
pept	5849	6059	tat protein, exon 2 (first expressed exon)
pept	8396	8486	tat protein, exon 3 (AA at 8397)
pept	5964	6059	rev protein, exon 2 (first expressed exon)
pept	8396	8667	rev protein, exon 3 (AA at 8398)
pept	6073	6144	vpu protein (premature termination)
pept	6239	8808	envelope polyprotein
pept	8811	9358	nef protein (premature termination at 9357 relative to other HIV-1 sequences)
pre-msg	454	9655	genomic mRNA
pre-msg	454	9655	tat, rev, nef subgenomic mRNA
IVS	740	5791	tat, rev, nef subgenomic mRNA intron 1
IVS	6060	8395	tat cds intron 2
IVS	6060	8395	rev cds intron 2
IVS	6060	8395	tat, rev, nef subgenomic mRNA intron 2
LTR	1	633	5' LTR
LTR	9106	9738	3' LTR
rpt	453	550	R repeat 5' copy
rpt	9558	9655	R repeat 3' copy
binding	376	395	Sp1 binding site III
binding	387	396	Sp1 binding site II
binding	398	407	Sp1 binding site I
binding	635	652	primer (Lys-tRNA) binding site
site	3783	3785	pol cds in-frame stop codon
signal	9631	9636	mRNA polyadenylation signal
BASE COUNT	3463 a	1763 c	2344 g 2162 t
ORIGIN	Left end of viral genome		

Initial Score = 596 Optimized Score = 2203 Significance = 0.00
 Residue Identity = 90% Matches = 2233 Mismatches = 170
 Gaps = 58 Conservative Substitutions = 0

```

X      10      20      30      40      50      60      70
ATGAGAGTGAAGGAGAAATATCAACACTTGTGGAGATGGGGGTGGAAATGGGGGCACCATGCTCCTTGGGATA
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
ATGAGAGTGAAGG-AGATCAGGAGGAATTAT-CAG-CACTGGTGGGGATGGGGGCACGATGCTCCTTGGGTTA
6240      6250      6260      6270      6280      6290      6300

      80      90     100     110     120     130     140
TTGATGATCTGTAGTGTCTACAGAAAAATTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCA
||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
TTAATGATCTGTAGTGTCTACAGAAAAATTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCA
6310      6320      6330      6340      6350      6360      6370

     150     160     170     180     190     200     210
ACCACCACCTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACAT
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
ACCACCACCTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACAA

```

6380 6390 6400 6410 6420 6430 6440 6450

220 230 240 250 260 270 280
GCCTGTGTACCCACACACCCCAACCCACACAGAGTAGTATTGGTAAATGTGACAGAAAATTTTAACATGTGG
|||||
GCCTGTGTACCCACACACCCCAACCCACACAGAGTAGAATTGGTAAATGTGACAGAAAATTTTAACATGTGG
6460 6470 6480 6490 6500 6510 6520

290 300 310 320 330 340 350 360
AAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTA
|||||
AAACATAACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTA
6530 6540 6550 6560 6570 6580 6590

370 380 390 400 410 420
AAATTAACCCCACTCTGTGTAGTTTAAAGTGCACCTGATTTGGGGAATGCTACTAATACCAATACTAG----
|||||
AAATTAACCCCACTCTGTGTAGTTTAAAGTGCACCTGATTTGGGGAATGCTACTAATACCAATACTAGTTACT
6600 6610 6620 6630 6640 6650 6660

430 440 450 460 470 480 490
--TAATACCAATAGTAGTAGCGGGGAATGATGATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATC
|||||
GCTAATAACAATAGTAATAGCGAGGGGAACAATAAAGG-GA--GGAGAAATGAAAACTGCTCTTTCAATATC
6670 6680 6690 6700 6710 6720 6730

500 510 520 530 540 550 560 570
AGCACAAGNATAAGAGGTAAAGGTGCAGAAAGAATATGCATTTTTTTATAAACTTGATATAATACCAATAGAT
|||||
ACCACAAGCATAAGAGGTAAAGGTGCAGAAAGAATATGCACCTTCTTTATAAACTTGATATAGTATCAATAGAT
6740 6750 6760 6770 6780 6790 6800

590 590 600 610 620 630 640
AATGATACTACCAAGCTATACGTTGACAGTTGTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCC
|||||
AATGATACTACCAAGCTATACGTTGATAAGTTGTAATACCTCAGTCATTACACAAGCTTGTCCAAAGATATCC
6810 6820 6830 6840 6850 6860 6870 6880

650 660 670 680 690 700 710
TTTGAGCCCAATTCCCAATACATTATTGTGCCCCGGCTGGTTTTTGCATTCTAAAATGTAATAATAAGACGTTT
|||||
TTTGAGCCCAATTCCCAATACATTATTGTGCCCCGGCTGGTTTTTGCATTCTAAAATGTAACGATAAAAAGTTT
6890 6900 6910 6920 6930 6940 6950

720 730 740 750 760 770 780
AATGGAAACAGGACCAATGTACAAATGTGAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACT
|||||
AGTGGAAACAGGACCAATGTACAAATGTGAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACT
6960 6970 6980 6990 7000 7010 7020

790 800 810 820 830 840 850
CAACTGCTGTTTGAATGGCAGTCTAGCAGAGAGAGAGGTAGTAATTAGATCTGCCAATTTACAGACAATGCT
|||||
CAACTGCTGTTTGAATGGCAGTCTAGCAGAGAGAGAGGTAGTAATTAGATCTGAGAAATTTCACTGATAATGCT
7030 7040 7050 7060 7070 7080 7090

860 870 880 890 900 910 920 930
AAAACCATATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAGACCCCAACAACAATACAAGAAAA
|||||
AAAACCATATAGTACATCTGAATGAATCTGTACAAATTAATTGTACAGACCCCAACTACAATAAAGAAAA
7100 7110 7120 7130 7140 7150 7160

940 950 960 970 980 990
AGTATCCGTATCCAGAGGGGAGCCAGGGAGAGSCATTTGTTACAATAGGAAA---AATAGGAAATATGAGACAA
|||||
AGGATACATAT---AG---GAGGAGGGAGAGSCATTTTATACAGCAAAAATATAATAGGAACTATAAGACAA

7170	7180	7190	7200	7210	7220	7230
------	------	------	------	------	------	------

1000	1010	1020	1030	1040	1050	1060	1070
GCACATTGTAACATTAGTAGAGCAAAATGCAATGCCACTTTAAACAGATAGCTAGCAAATTAAGAGAACAA							
GCACATTGTAACATTAGTAGAGCAAAATGCAATGCCACTTTAAGACAGATAGTTAGCAAATTAAGAGAACAA							
7240	7250	7260	7270	7280	7290	7300	

1080	1090	1100	1110	1120	1130	1140	
TTTGGAAATAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTT							
TTT---AAGAATAAAACAATAATCTTTAATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTT							
7310	7320	7330	7340	7350	7360	7370	

1150	1160	1170	1180	1190	1200	1210	
AATTGTGGAGGGGAATTTTTCTACTGTAATTCAACACAACCTGTTTAATAGTACTTGG---TTTAATAGTACT							
AATTGTGGAGGGGAATTTTTCTACTGTAATACATCACCCTGTTTAATAGTACTTGGGAATGGTAATAATACT							
7380	7390	7400	7410	7420	7430	7440	

1220	1230	1240	1250	1260	1270	1280	
TGG---AGTACTGAAGGGTCAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCAGAATAAAACAA							
TGGAAATAACTACAGGGTCAATAACAAT-----ATCACACTTCAATGCAAATAAAACAA							
7450	7460	7470		7480	7490	7500	

1290	1300	1310	1320	1330	1340	1350	
TTTATAAACATGTGSCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATCAGCGGACAAATTAGATGTTCA							
ATTATAAACATGTGSCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATGAGGACAAATTAGATGTTCA							
7510	7520	7530	7540	7550	7560	7570	

1360	1370	1380	1390	1400	1410		
TCAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAACA---AC---AATGGGTCCGAGATCTTC							
TCAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAAGGACACGGACACGAACGACACCGAGATCTTC							
7580	7590	7600	7610	7620	7630	7640	

1420	1430	1440	1450	1460	1470	1480	1490
AGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAACAATTGAA							
AGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAACAATTGAA							
7650	7660	7670	7680	7690	7700	7710	7720

1500	1510	1520	1530	1540	1550	1560	
CCATTAGGAGTAGCACCCACCAAGGCAAGAGAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGA							
CCATTAGGAGTAGCACCCACCAAGGCAAGAGAGAGTGGTGCAGAGAGAAAAAGAGCAGCG---ATAGGA							
7730	7740	7750	7760	7770	7780		

1570	1580	1590	1600	1610	1620	1630	
GCTTTGTTCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAG							
GCTCTGTTCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAG							
7790	7800	7810	7820	7830	7840	7850	7860

1640	1650	1660	1670	1680	1690	1700	
GCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCAT							
GCCAGACTATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCAT							
7870	7880	7890	7900	7910	7920	7930	

1710	1720	1730	1740	1750	1760	1770	
CTGTTGCAACTCACAGTCTGGGGCATCAGCAGCTCCAGGCAAGAGTCTGCTGTGGAAAGATACCTAAG							
ATGTTGCAACTCACAGTCTGGGGCATCAGCAGCTCCAGGCAAGAGTCTGCTGTGGAAAGATACCTAAG							

7840 7850 7860 7870 7880 7890 8000

1780 1790 1800 1810 1820 1830 1840 1850
GATCAACAGCTCCTGGGGATTGTTGGGTTGCTCTGGAAACTCATTGTCACCACTGCTGTGCCTTGGGAATGCT
|||||
GATCAACAGCTCCTGGGGTTTGGGTTGCTCTGGAAACTCATTGTCACCACTACTGTGCCTTGGGAATGCT
8010 8020 8030 8040 8050 8060 8070

1860 1870 1880 1890 1900 1910 1920
AGTTGGAGTAATAAACTCTGGAACAGATTTGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAAC
|||||
AGTTGGAGTAATAAATCTCTGGATGATATTTGGAATAACATGACCTGGATGCAGTGGGAAAGAGAAATTGAC
8080 8090 8100 8110 8120 8130 8140

1930 1940 1950 1960 1970 1980 1990
AATTACACAAGCTTAATACATTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTA
|||||
AATTACACAAGCTTAATATACTCATTACTAGAAAAATCGCAAAACCAACAAGAAAAGAATGAACAAGAATTA
8150 8160 8170 8180 8190 8200 8210 8220

2000 2010 2020 2030 2040 2050 2060
TTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGTTTAACATAACAAATTGGCTGTGGTATATAAAAATA
|||||
TTGGAATTGGATAAATGGGCAAGTTTGTGGAATTGTTTGACATAACAAATTGGCTGTGGTATATAAAAATA
8230 8240 8250 8260 8270 8280 8290

2070 2080 2090 2100 2110 2120 2130
TTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGGCTGTACTTTCTATAGTGAATAGAGTT
|||||
TTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGGCTGTACTTTCTATAGTGAATAGAGTT
8300 8310 8320 8330 8340 8350 8360

2140 2150 2160 2170 2180 2190 2200 2210
AGGCAGGGATATTCACCATTTATCGTTTCAGACCCACCTCCCAACCCGAGGGGACCCGACAGGCCCGAAGGA
|||||
AGGCAGGGATATTCACCATTTATCGTTTCAGACCCGCCCCCAGTTCCGAGGGGACCCGACAGGCCCGAAGGA
8370 8380 8390 8400 8410 8420 8430

2220 2230 2240 2250 2260 2270 2280
ATAGAAGAGAGAGGTGGAGAGAGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTATC
|||||
ATCGAAGAGAGAGGTGGAGAGAGAGACAGAGACACATCCGGTGGATTAGTGCATGGATTCTTAGCAATTATC
8440 8450 8460 8470 8480 8490 8500

2290 2300 2310 2320 2330 2340 2350
TGGCAGGATCTCCGAGCCTTGTCCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGAG
|||||
TGGGTGACCTCCGAGCC-TGTTCTCTTCAGCTACCACCAC---AGAGACTTACTCTTGATTGCAGCGAG
8510 8520 8530 8540 8550 8560 8570

2360 2370 2380 2390 2400 2410 2420
GATTGTGGAACCTTCTGGGACGCAAGGGGTGGGAAGCCCTCAAATATTGGTGGAACTCTCTACAGTATTGGAG
|||||
GATTGTGGAACCTTCTGGGACGCAAGGGGTGGGAAGTCTCAAATATTGGTGGAACTCTCTACAGTATTGGAG
8580 8590 8600 8610 8620 8630 8640

2430 2440
TCAGGAACCTAAAG
|||||
TCAGGAACCTAAAG
8650 8660 X

LOCUS HIVCDC42 3373 bp ss-RNA VRL 15-JUN-1989
 DEFINITION Human immunodeficiency virus type 1, isolate CDC-451, tat, rev, env and nef genes.
 ACCESSION M13137
 KEYWORDS env gene; tat gene.
 SEGMENT 2 of 2
 SOURCE Human Immunodeficiency virus type 1 (HIV-1), isolate CDC-451, unintegrated circular proviral DNA.
 ORGANISM Human Immunodeficiency virus type 1
 Viridae; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.
 REFERENCE 1 (bases 1 to 3373)
 AUTHORS Desai, S. M., Kalyanaraman, V. S., Casey, J. M., Srinivasan, A., Andersen, P. F. and Devare, S. G.
 TITLE Molecular cloning and primary nucleotide sequence analysis of a distinct human immunodeficiency virus isolate reveal significant divergence in its genomic sequences
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 83, 8380-8384 (1986)
 STANDARD full staff_review
 COMMENT Kindly submitted in computer-readable form by [1]. The normal start codon for the nef gene is not present; the ATG at 3142 may serve this role.

FEATURES	from	to/span	description
pept	1	52	vpr protein, partial (AA at 2)
pept	88	302	tat protein, exon 2 (first expressed exon)
	2667	2757	tat protein, exon 3 (AA at 2668)
pept	227	302	rev protein, exon 2 (first expressed exon)
	2667	2941	rev protein, exon 3 (AA at 2669)
pept, ps	319	555	vpu protein (in-frame stops at bases 451 and 484)
pept	477	3063	envelope polyprotein
pre-msg	1	3373	genomic mRNA
pre-msg	1	3373	tat, rev, nef subgenomic mRNA
IVS	1	34	tat, rev, nef subgenomic mRNA intron 1
IVS	303	2666	tat cds intron 2
IVS	303	2666	rev cds intron 2
IVS	303	2666	tat, rev, nef subgenomic mRNA intron 2
BASE COUNT	1174 a	585 c	808 g 805 t
ORIGIN	1 bp upstream of EcoRI site; about 3.6 kb after segment 1.		

Initial Score = 932 Optimized Score = 2191 Significance = 0.00
 Residue Identity = 89% Matches = 2223 Mismatches = 194
 Gaps = 60 Conservative Substitutions = 0

```

X      10      20      30      40      50      60      70
ATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAATGGGGCACCATGCTCCTTGGGATA
||||| ||||| ||| | | | | |||| | | ||||| ||||| ||||| ||||| ||||| ||||| |||||
ATGAGAGCGAAGG-GGATCAGGAAGAATTGT-CAG-CACTTGTGGAGATGGGGCACCATGCTCCTTGGGAATG
X      430     500     510     520     530     540     550

      80      90     100     110     120     130     140
TTGATGATCTGTAGTCTACAGAAAAATTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCA
||||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
TTGATGATCTGTAGTCTGCAGCAAACTTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCA
      560     570     580     590     600     610     620

     150     160     170     180     190     200     210
ACCACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACAT
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
ACCACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGCACATAATGTTTGGGCCACACAT
     630     640     650     660     670     680     690

     220     230     240     250     260     270     280
GCCGTGTGTACCCACAGACCCCAACCCACAGAAGTAGTATTGGTAAATGTGACAGAAAATTTTAACATGTGG
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
GCCGTGTGTACCCACAGAACCCCAACCCACAGAAGTAGTATTGGTAAATGTGACAGAAAATTTTAACATGTGG

```

700 710 720 730 740 750 760

290 300 310 320 330 340 350 360
AAAAATGACATCGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCCTAAAGCCCATGTGTA
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
AAAATAAOCATCGTAGAACAGATGCATGAGGATATAATCAGCTTATGGGATCAAAGCCCTAAAGCCCATGTGTA
770 780 790 800 810 820 830

 370 380 390 400 410 420 430
AAATTAAACCCCACACTCTGTGTTAGTTTTAAAGTGCACTGATTTGGGGAATGCTACTAATACCAATACTA--GTA
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
AAACTAACCCCACACTCTGTGTTACTTTAAATTGCACTGATTTGAATACTAATAATACTACTAATACTACTGAA
840 850 860 870 880 890 900 910

 440 450 460 470 480 490 500
ATACCAATAGTAGTAGCGGGGAAATGATGATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCA
|| | ||||| ||||| || || | || | ||||| || || | ||||| || || | ||||| || || | ||||| || || | ||||| || || |
CTATCAATAATAGTAGTTTTGGGAACAACG--GGGTAAAGGAGAAATGAGAACTGTTCTTTCAATATCACCA
920 930 940 950 960 970 980

 510 520 530 540 550 560
CAASNATAAGAGGTAAAGTGACAGAAGGAATATGCATTTTTTTTATAAACTTGATATAATACCAATAG-----
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CAAGCATAAGAWATAAGGTGCAGAGAGGAATATGCATTGTTTTATAAACTTGATGTAGAACCAATAGATGATA
990 1000 1010 1020 1030 1040 1050

570 580 590 600 610 620 630
ATAATGATACTACCAGC-----TATACGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCCTGTC
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
ATAAAAATACTACCACACACACCAATA TAGGTTGATAAATTGTAACACCTCAGTCATTACACAGGCCCTGTC
1060 1070 1080 1090 1100 1110 1120

 640 650 660 670 680 690 700
CAAAGGTATCCT TTGAGCCAATTCCCATACATTATTGTGCCCGGCTGGTTTTGCGATTCTAAATGTAATA
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTACCCCGACTGGTTTTGCACTTCTAAAGTGTAAAG
1130 1140 1150 1160 1170 1180 1190

 710 720 730 740 750 760 770
ATAAGACGTTCAATGGAACAGGACCATGTACAAATGTGAGCACAGTACAATGTACACATGGAATTAGGCCAG
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
ATAAGAACGTTCAATGGAGACAGGACCATGTACAAATGTGAGCACAGTACAATGTACACATGGAATTAGGCCAG
1200 1210 1220 1230 1240 1250 1260

 780 790 800 810 820 830 840
TAGTATCAACTCAACTGCTGTTGAATGGCAGTCTAGCAGAAGAAGAGGGTAGTAATTAGATCTGCCAATTTCA
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
TAGTGTCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGGGTAGTAATTAGATCTGAAAATTTCA
1270 1280 1290 1300 1310 1320 1330 1340

 850 860 870 880 890 900 910
CAGACAATGCTAAAACCATATAATGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACA
| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CGAACAATGCTAAAACCATATAATGTACAGCTGAATGTATCTGTAGAAATTAATTGTACAAGACCCAACAACC
1350 1360 1370 1380 1390 1400 1410

920 930 940 950 960 970 980
ATACAGANAAAGTATCCGTATCCAGAGGGGACCAGGGAGAGCATTGTGTTACAATAGGAAAAA---TAGGAA
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
ATACAGANAAAG----GGTAAC--GCTAGGACCAGGGAGAGTATGGTATACAACAGGAGAAATACTAGGAA
1420 1430 1440 1450 1460 1470

990 1000 1010 1020 1030 1040 1050 1060
ATATGAGACAACACATTGTAACTTTAGTAGAGCAAAATGCAATGCCACTTTAAACAGATAGCTAGCAAT
|||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
ATATGAGACAACACATTGTAACTTTAGTAGAGCAAAATGGAATAACACTTTACAACAGATAGCTAGCAACT

[illegible]

2270 2280 2290 2300 2310 2320 2330

1840 1850 1860 1870 1880 1890 1900
TGTGCCTTGGAAATGCTAGTTTGGAGTAATAAATCTCTGGAACAGATTTTGAATAACATGACCTGGATGGAGTG
|||||
TGTGCCTTGGAAATGCTAGTTTGGAGTAATAAATCTCTGGATCAGATTTTGAATAACATGACCTGGATGGAGTG
2340 2350 2360 2370 2380 2390 2400

1910 1920 1930 1940 1950 1960 1970
GGACAGAGAAATTAAACAATTACACAGGCTTAATACATTCCCTTAATTGAAGAATCGCAAACACAGCAAGAAAA
|||||
GGACAGAGAAATTAAACAATTACACAGGCTTAATATACACTTTAATTGAAGAATCGCAAACCAACAAGAAAA
2410 2420 2430 2440 2450 2460 2470

1980 1990 2000 2010 2020 2030 2040 2050
GAATGAACAGGAAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAAATTGGTTTAACATAACAAATTGGCT
|||||
GAATCAACAGGAACTATTGCAATTAGATAAGTGGGCAAGTTTGTGGACTTGGTCTGACATAACAAATGGCT
2480 2490 2500 2510 2520 2530 2540

2060 2070 2080 2090 2100 2110 2120
GTGGTATATAAAATATTTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTC
|||||
GTGGTATATAAAATATTTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTGCTTTC
2550 2560 2570 2580 2590 2600 2610 2620

2130 2140 2150 2160 2170 2180 2190
TATAGTGAATACAGTTAGGCAGGATATTCACCATTATCGTTTCAGACCCACCTCCCAACCCCGAGGGGACC
|||||
TATAGTGAATACAGTTAGGCAGGATATTCACCATTATCGTTTCAGACCCCTCCTCCCAACCCCGAGGGGACC
2630 2640 2650 2660 2670 2680 2690

2200 2210 2220 2230 2240 2250 2260
CGACAGGCCCGAAGGAATAGAGGAAGAGGTGGAGAGAGAGACAGACAGATCCATTGATTAGTGAACGG
|||||
CGACAGGCCCGAAGGAACCGAAGAGAGGAGGTGGAGAGAGAGAGGCAGAGACGGATCCACTGATTAGTGCATGG
2700 2710 2720 2730 2740 2750 2760

2270 2280 2290 2300 2310 2320 2330
ATCCTTAGCAGCTTATCTGGGACGATCTGCGGAGCCTTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTAC
|||||
CTTCTTAGCAGCTTATCTGGGACGATCTGCGGAGCC-TGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTAC
2770 2780 2790 2800 2810 2820 2830

2340 2350 2360 2370 2380 2390 2400 2410
TCTTGATTGTACGAGGATTGTGGAACCTCTGGGACGCAGGGGGTGGGAAGCCCTCAAATATTGGTGGAAATC
|||||
TCTTGATTGTACGAGGATTGTGGAACCTCTGGGACGCAGGGGGTGGGAAGTCCTCAAATATTGGTGGAAATC
2840 2850 2860 2870 2880 2890 2900

2420 2430 2440
TCCTACAGTATTGGAGTCAGGAAGTAAG
|||||
TCCTACAGTATTGGAGTCAGGAAGTAAG
2910 2920 2930 X

4. KUNZ-159-CL33. SEQ

HIVPV22 Human immunodeficiency virus type 1, isolate PV22,

LOCUS HIVPV22 8770 bp ss-RNA VRL 15-JUN-1989
DEFINITION Human immunodeficiency virus type 1, isolate PV22, complete genome
(HSV/HTLV-III proviral DNA).
ACCESSION K02093
KEYWORDS TAP: protein; acquired immune deficiency syndrome; complete genome;
env: gene; gag: gene; long terminal repeat; pol: gene; pol: protein;

provirus; rev gene; reverse transcriptase; tat gene; trans-activator.

SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate PV22 (from H9-derived family), proviral DNA.

ORGANISM Human immunodeficiency virus type 1
Virusidae; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.

REFERENCE 1 (bases 1 to 8770; revised sequence, personal communication)

AUTHORS Muesing, M. A. , Smith, D. H. , Cabradilla, C. D. , Benton, C. V. , Lasky, L. A. and Capon, D. J.

TITLE Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus

JOURNAL Nature 313, 450-458 (1985)

STANDARD full staff_review

REFERENCE 2 (bases 2111 to 2112; revises [1])

AUTHORS Muesing, M. A.

JOURNAL Unpublished (1987) Whitehead Inst Cambridge, Mass

STANDARD full staff_review

COMMENT This sequence for a H9/HTLV-III virus was determined from one complete proviral clone [1]. Additionally, several cDNA clones of the viral RNA were sequenced for comparison with the entire proviral sequence. The differences between cDNA and proviral DNA are extensive and are listed in the Sites Table as variations. The authors believe that the variations may be due in part to different strains in the H9/HTLV-III cell line, because it was established by infection with material from several AIDS patients. With the addition of g at 2111, gag cds and pol cds are very close to those of HXB2, BRU, and related HIV viruses. For details and other references pertaining to Sites and Features, see the HIV reference entry.

FEATURES	from	to/span	description
pept	758	2337	gag polyprotein precursor
pept	(2084	5141	pol polyprotein (NH2-terminus uncertain; AA at 2084)
pept	5086	5664	vif protein
pept	5604	5840	vpr protein
pept	5876	6090	tat protein, exon 2 (first expressed exon)
	8421	8466	tat protein, exon 3 (AA at 8422)
pept	6015	6090	rev protein, exon 2 (first expressed exon)
	8421	8685	rev protein, exon 3 (AA at 8423)
pept	6107	6352	vpu protein
pept	6267	8837	envelope polyprotein
pept	8838	9459	nef protein
pre-msg	464	5570	genomic mRNA
pre-msg	464	5678	tat, rev, nef subgenomic mRNA
IVS	753	5322	tat, rev, nef subgenomic mRNA intron 1
IVS	6081	8420	tat cds intron 2
IVS	6081	8420	rev cds intron 2
IVS	6081	8420	tat, rev, nef subgenomic mRNA intron 2
LTR	10	643	5' LTR
LTR	9128	9761	3' LTR
rpt	463	560	R repeat 5' copy
rpt	9581	9678	R repeat 3' copy
binding	386	395	Sp1 binding site III
binding	397	406	Sp1 binding site II
binding	406	417	Sp1 binding site I
binding	645	662	primer (Lys-tRNA) binding site
variant	510	510	a in provirus; g in cDNA [1]
variant	575	575	g in provirus; a in cDNA [1]
revision	2111	2112	gg in [2]; g in [1]
variant	5716	5716	g in provirus; a in cDNA [1]
variant	5992	5992	a in provirus; g in cDNA [1]
variant	6007	6007	c in provirus; t in cDNA [1]
variant	6047	6047	c in provirus; g in cDNA [1]
variant	6051	6051	c in provirus; a in cDNA [1]
variant	6055	6057	agg in provirus; gaa in cDNA [1]

[illegible]

[illegible]

SEGMENT 2 of 2
SOURCE Human immunodeficiency virus type 1 (HIV-1) proviral DNA clone
2110.

REFERENCE 1 (Pages 1 to 3032)

TITLE Complete nucleotide sequence of the AIDS virus, HTLV-III

JOURNAL *Nature* 313, 277-284 (1985)

STANDARD full staff review

FEATURES	from	to/stop	description
pep4	111	1530	gag polyprotein precursor
pept	1107	4053	pol polyprotein (NH2-terminus uncertain; AA at 1407)
pep6	4393	4397	vif protein
pep7	4917	5153	vpr protein
pep8	5135	5303	tat protein, exon 2 (first expressed exon)
	7735	7739	tat protein, exon 3 (AA at 7735)
pep9	5343	5403	rev protein, exon 2 (first expressed exon)
	7734	5403	rev protein, exon 3 (AA at 7736)
pep10	5420	5555	vpu protein
pep11	5555	6100	envelope polyprotein
pep12	6151	6513	nef protein, exon 3 (first expressed exon; premature termination)
prom1ag (3532	genomic mRNA
prom2ag (3532	tat, rev, nef subgenomic mRNA
IVS1	55	5155	tat, rev, nef subgenomic mRNA intron 1
IVS2	5405	7733	tat cds intron 2
IVS3	5405	7733	rev cds intron 2
IVS4	5405	7733	tat, rev, nef subgenomic mRNA intron 2
LTR	6441	3532	3' LTR
rep1	5555	3532	R repeat 3' copy
BASE COUNT	3597 a	1312 c	2147 g 1976 t
ORIGIN	about 120 bp downstream from HIVBH101.		

Residence Telephone: _____; Cell Phone: _____; E-mail: _____

[illegible]

6270 6280 6290

740 750 760 770 780 790 800
CAGATGTGACGACGATACACATGATCAGCATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCA
|||||
CAAAATGTGACGACGATACACATGATCAGCATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCA
6300 6310 6320 6330 6340 6350 6360

810 820 830 840 850 860 870
GTCTAGCAGAGAGAGAGGTACTTATAGATCTGCCAATTTACAGACAATGCTAAAACCATAATAGTACAGC
|||||
GTCTAGCAGAGAGAGAGGTACTTATAGATCTGCCAATTTACAGACAATGCTAAAACCATAATAGTACAGC
6370 6380 6390 6400 6410 6420 6430 6440

880 890 900 910 920 930 940
TGAACCAATCTGAGAAATTATGTTACAGAGACCCACACAACAATACAAGAAAAAGTATCCGTATCCAGAGGG
|||||
TGAACCAATCTGAGAAATTATGTTACAGAGACCCACACAACAATACAAGAAAAAGTATCCGTATCCAGAGAG
6450 6460 6470 6480 6490 6500 6510

950 960 970 980 990 1000 1010 1020
GACCAAGGAGAGGCAATTTGTTACAAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAG
|||||
GACCAAGGAGAGGCAATTTGTTACAAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAG
6520 6530 6540 6550 6560 6570 6580

1030 1040 1050 1060 1070 1080 1090
CAAAATGCAATCCCAATTTAGAGAGAGATAGCTAGCAATTAAGAGAACAAATTTGGAAATAATAAAACAATAA
|||||
CAAAATGCAATCCCAATTTAGAGAGAGATAGCTAGCAATTAAGAGAACAAATTTGGAAATAATAAAACAATAA
6590 6600 6610 6620 6630 6640 6650

1100 1110 1120 1130 1140 1150 1160
TCTTTAGCAATCCTCAGGAGGAGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTTCT
|||||
TCTTTAGCAATCCTCAGGAGGAGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTTCT
6660 6670 6680 6690 6700 6710 6720

1170 1180 1190 1200 1210 1220 1230
ACTGTAAATTCAGACAACTGTTTATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACA
|||||
ACTGTAAATTCAGACAACTGTTTATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACA
6730 6740 6750 6760 6770 6780 6790 6800

1240 1250 1260 1270 1280 1290 1300
CTGAAGGAAGTACAGCAATCAGCCTCCCATGCAGAAATAAAACAATTTATAAACATGTGGCAGGAAGTAGGAA
|||||
CTGAAGGAAGTACAGCAATCAGCCTCCCATGCAGAAATAAAACAATTTATAAACATGTGGCAGGAAGTAGGAA
6810 6820 6830 6840 6850 6860 6870

1310 1320 1330 1340 1350 1360 1370 1380
AAGCAATGTATGCCCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAAATATTACAGGGCTGCTATTAACAA
|||||
AAGCAATGTATGCCCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAAATATTACAGGGCTGCTATTAACAA
6880 6890 6900 6910 6920 6930 6940

1390 1400 1410 1420 1430 1440 1450
GAGATGGTGGTAATACAGCAATCGGTCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGA
|||||
GAGATGGTGGTAATACAGCAATCGGTCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGA
6950 6960 6970 6980 6990 7000 7010

1460 1470 1480 1490 1500 1510 1520
GAAGTGAATTATATAAATATAAGGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAGAGAA
|||||
GAAGTGAATTATATAAATATAAGGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAGAGAA

[illegible]

7810	7820	7830	7840	7850	7860	7870	7880
2320	2330	2340	2350	2360	2370	2380	
TACCTTGGATTGCTGACACTTACTCTTGATTGTAAACBAGGATTGTGGAACTTCTGGGACGCAGGGGGGTGGGAA							
TACCTTGGATTGCTGACACTTACTCTTGATTGTAAACBAGGATTGTGGAACTTCTGGGACGCAGGGGGGTGGGAA							
7890	7900	7910	7920	7930	7940	7950	
2390	2400	2410	2420	2430	2440		
GGGCTGCAATATCTCTGGAACTCTGTACAGTATTGAGTCAGGAACTAAAG							
GGGCTGCAATATCTCTGGAACTCTGTACAGTATTGAGTCAGGAGCTAAAG							
7960	7970	7980	7990	8000	X		

[illegible]

[illegible]

2070	2080	2090	2100	2110	2120	2130
1710	1720	1730	1740	1750	1760	1770
AATTTCGCTGAGGAGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAG						
AATTTCGCTGAGGAGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAG						
2140	2150	2160	2170	2180	2190	2200
1780	1790	1800	1810	1820	1830	1840
GCAAGGATCTCTGCTGATGGAAGGATACCTAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAA						
GCAAGGATCTCTGCTGATGGAAGGATACCTAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAA						
2210	2220	2230	2240	2250	2260	2270
1820	1830	1840	1850	1860	1870	1880
CTGATTTGCAACCACTGCTGTGCTTGGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGGAATAAC						
CTGATTTGCAACCACTGCTGTGCTTGGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGGAATCAC						
2280	2290	2300	2310	2320	2330	2340
1900	1910	1920	1930	1940	1950	1960
ATGACCTGAGATGAGTGGGACAGAGAAATTACGAATTACACAAGCTTAATACATTCCTTAATTGAAGAATCG						
ACGACGTGGATGAGTGGGACAGAGAAATTACGAATTACACAAGCTTAATACATTCCTTAATTGAAGAATCG						
2350	2360	2370	2380	2390	2400	2410
1970	1980	1990	2000	2010	2020	2030
CAAAACCAAGCAAGAAAGAAATGAACAAGCAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTT						
CAAAACCAAGCAAGAAAGAAATGAACAAGCAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTT						
2430	2440	2450	2460	2470	2480	2490
2040	2050	2060	2070	2080	2090	2100
AACAATAACAAATGGGCTGTGATATATAAAATATTTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATA						
AACAATAACAAATGGGCTGTGATATATAAAATATTTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATA						
2500	2510	2520	2530	2540	2550	2560
2110	2120	2130	2140	2150	2160	2170
GTTTTTGTCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTC						
GTTTTTGTCTGTACTTTCTGTAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTC						
2570	2580	2590	2600	2610	2620	2630
2180	2190	2200	2210	2220	2230	2240
CCACCCCGAGGGGAGCCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCC						
CCACCCCGAGGGGAGCCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCC						
2630	2640	2650	2660	2670	2680	2690
2250	2260	2270	2280	2290	2300	2310
ATTGATTAAGTGAACGATGCTTAGCACTTATCTGGGACGATCTGCGGAGCCTTGTGCCTCTTCAGCTACCA						
ATTGATTAAGTGAACGATGCTTAGCACTTATCTGGGACGATCTGCGGAGCCTTGTGCCTCTTCAGCTACCA						
2720	2730	2740	2750	2760	2770	2780
2330	2340	2350	2360	2370	2380	2390
CCGCTTGAGAGAGCTTACTCTTGATTTGTAACGAGGATTGTGGAACCTTCTGGGACGACAGGGGGTGGGAAGCCCT						
CCGCTTGAGAGAGCTTACTCTTGATTTGTAACGAGGATTGTGGAACCTTCTGGGACGACAGGGGGTGGGAAGCCCT						
2790	2800	2810	2820	2830	2840	2850
2400	2410	2420	2430	X		
CAGATATTGCTGGAATCTCTTACAGTATTGGAGTCAGGAACTAAG						
CAGATATTGCTGGAATCTCTTACAGTATTGGAGTCAGGAACTAAG						

7. KUNZ-155-7L33.520

HIVNL43 Human immunodeficiency virus type 1, NY5/BRU (LAV-

LOCUS HIVNL43 5709 bp ss-RNA VRL 15-JUN-1989

DEFINITION Human immunodeficiency virus type 1, NY5/BRU (LAV-1) recombinant clone pNL4-3.

ACCESSION M18421

KEYWORDS .

SOURCE Human immunodeficiency virus type 1 (HIV-1), NY5/BRU (LAV-1) recombinant clone pNL4-3.

ORGANISM Human immunodeficiency virus type 1
 Viridae; ss-RNA enveloped viruses; Retroviridae;
 Lentivirinae.

REFERENCE 1 (bases 1 to 5709)

AUTHORS Buckler,C.E. , Buckler-White,A.J. , Willey,R.L. and McCoy,J.

JOURNAL Unpublished (1988).

STANDARD full staff_review

REFERENCE 2 (bases 1 to 5709)

AUTHORS Adechi,A. , Gendelman,H.E. , Koenig,S. , Folks,T. , Willey,R. , Roberson,A. and Martin,M.A.

TITLE Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone

JOURNAL J. Virol. 53, 284-291 (1986)

STANDARD full staff_review

REFERENCE 3 (sites; revisions of [1])

AUTHORS Buckler,C.E.

JOURNAL Unpublished (1989)

STANDARD full staff_review

COMMENT Clean copy of sequence [1] kindly provided by Chuck Buckler, NIAID, Bethesda, MD, 24-JUN-1988. The construction of pNL4-3 has been described in [2]. pNL4-3 is a recombinant (infectious) proviral clone that contains DNA from HIV isolates NY5 (5' half) and BRU (3' half). The site of recombination is the EcoRI site at positions 5743-5748.

The length and sequence of the vpr coding region corresponds to that of the BRU, SC, SF2, MAL and ELI isolates. The vpr coding region of these isolates is about 18 amino acid residues longer than the vpr coding region of the IIIb isolates. In HIVNL43, this shift is due to a single base deletion (with respect to the IIIb's) at position 5770. The sequence at this position is "atttc" in HIVNL43 and "atttcc" in HIVHXB2.

The original BRU clone, sequenced by Wain-Hobson, et al. (Cell 40, 9-17 (1985)), and the BRU portion of the pNL4-3 recombinant clone are different clones from the same BRU isolate.

Two of the revisions reported in the FEATURES produced changes in amino acid sequences. The revision at position 2421 changes one amino acid residue from 'R' to 'G' in the pol coding region. The revision at positions 8995-9000 changes three amino acid residues from 'AHT' to 'VTP' in the nef coding region.

FEATURES	from	to/span	description
pept	730	2292	gag polyprotein
pept	(2085	5095	pol polyprotein (NH2-terminus uncertain; AA at 2085)
pept	5041	5619	vif protein
pept	5559	5949	vpr protein
pept	5830	6044	tat protein, exon 2 (first expressed exon)
	8369	8414	tat protein, exon 3 (AA at 8370)
pept	5859	6044	rev protein, exon 2 (first expressed exon)
	8369	8543	rev protein, exon 3 (AA at 8371)
pept	6061	6306	vpu protein
pept	6221	8785	envelope polyprotein
pept	8787	9407	nef protein

pre-msg	455	5626	genomic mRNA
pre-msg	455	5626	tat, rev, nef subgenomic mRNA
IVS	744	5776	tat, rev, nef mRNA intron 1
IVS	6045	6368	tat cds intron 2
IVS	6045	6368	rev cds intron 2
IVS	6045	6368	tat, rev, nef mRNA intron 2
LTR	1	634	5' LTR
LTR	9076	9709	3' LTR
rpt	454	550	R repeat 5' copy
rpt	9529	9526	R repeat 3' copy
binding	377	386	Spl binding site III
binding	388	397	Spl binding site II
binding	399	408	Spl binding site I
binding	636	653	primer (Lys-tRNA) binding site
site	5743	5748	EcoRI site of recombination
recomb	5743	5744	HIV-1 isolate NY5 DNA end/HIV-1 isolate LAV DNA start
revision	182	183	at in [3]; tg in [1]
revision	194	194	g in [3]; c in [1]
revision	2421	2421	g in [3]; a in [1]
revision	8995	9000	tcacac in [3]; ctcaca in [1]
revision	9415	9415	c in [3]; a in [1]
signal	9602	9607	mRNA polyadenylation signal
BASE COUNT	3421 a	1756 c	2366 g 2166 t
ORIGIN	5' terminus of NY5 LTR		

400 410 420 430 440 450
TGCACCTGATTTGG-----GGAATGCTACTAAT---ACCAATACTAGTAATACCAATAGTAGTAGCGGGGAAATG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CACTCTGTGTTAGTTTTAAAGTGCG-ACTGATTTGAAGAATGATACTAATACCAATAGTAGTAGCGGGGAGAATG
6500 6510 6520 6530 6540 6550 6560

460 470 480 490 500 510 520
ATEATGGAGAAAAGGAGAGATAAAAAACTGCTCTTTCAATATCAGCACAAAGNATAAGAGGTAAAGGTGCAGAAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ATAATGGAGAAAAGGAGAGATAAAAAACTGCTCTTTCAATATCAGCACAAAGCATAGAGATAAGGTGCAGAAA
6670 6680 6690 6700 6710 6720 6730

530 540 550 560 570 580 590 600
GAAATATGCATTITTTTTATAAACTTGTATATAATACCAATAGATAATGATACTACCAGCTATACGTTGACAAGT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GAATATGCATTCTTTTATAAACTTGTATATAGTACCAATAGATAA-----TACCAGCTATAGGTTGATAAGT
6740 6750 6760 6770 6780 6790

610 620 630 640 650 660 670
TGTAACACCTTCAGTCATTACACAGGCCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCACATTATTGTGCC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TGTAACACCTTCAGTCATTACACAGGCCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCACATTATTGTGCC
6800 6810 6820 6830 6840 6850 6860 6870

680 690 700 710 720 730 740
CCGCTCGGTTTTTGCATTTCTAAATGTATAATAAGACGTTCAATGGAACAGGACCATGTACAAATGTCAGC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CCGCTCGGTTTTTGCATTTCTAAATGTATAATAAGACGTTCAATGGAACAGGACCATGTACAAATGTCAGC
6880 6890 6900 6910 6920 6930 6940

750 760 770 780 790 800 810
ACAGTACAATGTACACATGGAAATAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCAGTCTAGCAGAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACAGTACAATGTACACATGGAAATAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAA
6950 6960 6970 6980 6990 7000 7010

820 830 840 850 860 870 880
GAAGAGGTAGTAATTAGATCTGCCAATTTACACAGACAATGCTAAAACCATAATAGTACAGCTGAACCAATCT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GAAGATGTAGTAATTAGATCTGCCAATTTACACAGACAATGCTAAAACCATAATAGTACAGCTGAACACATCT
7020 7030 7040 7050 7060 7070 7080

890 900 910 920 930 940 950 960
GTAGAAATTAATTGTACAAAGACCCAACACAAATACAAGAAAAAGTATCCGTATCCAGAGGGGACCAGGGAGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GTAGAAATTAATTGTACAAAGACCCAACACAAATACAAGAAAAAGTATCCGTATCCAGAGGGGACCAGGGAGA
7090 7100 7110 7120 7130 7140 7150

970 980 990 1000 1010 1020 1030
GCAITTTGTTACAAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAGCAAATGCAAT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCAITTTGTTACAAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAGCAAATGGAAT
7160 7170 7180 7190 7200 7210 7220 7230

1040 1050 1060 1070 1080 1090 1100
GCCACTTTAAACAGATAGCTAGCAAATTAAGAGAACAAATTTGGAAATAATAAAACAATAATCTTTAAGCAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCCACTTTAAACAGATAGCTAGCAAATTAAGAGAACAAATTTGGAAATAATAAAACAATAATCTTTAAGCAA
7240 7250 7260 7270 7280 7290 7300

1110 1120 1130 1140 1150 1160 1170
TCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTCTACTGTAATTCA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTCTACTGTAATTCA
7310 7320 7330 7340 7350 7360 7370

1150	1150	1200	1210	1220	1230	1240
ACACAACCTGTTTAAATAGTACTTTGGTTTAAATAGTACTTTGGAGTACTGAAGGGTCAAATAACACTGAAGGAAGT						
ACACAACCTGTTTAAATAGTACTTTGGTTTAAATAGTACTTTGGAGTACTGAAGGGTCAAATAACACTGAAGGAAGT						
7330	7350	7400	7410	7420	7430	7440

1250	1250	1270	1280	1290	1300	1310	1320
GACACAATCAGACTCCCATGCGAATAAAACAATTTATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTAT							
GACACAATCAGACTCCCATGCGAATAAAACAATTTATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTAT							
7450	7460	7470	7490	7490	7500	7510	

	1330	1340	1350	1360	1370	1380	1390
GCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAAATATTACAGGGCTGCTATTAAACAGAGATGGTGGT							
GCCCTCCCATCAGTGGACAAATTAGATGTTTCATCAAATATTACTGGGCTGCTATTAAACAGAGATGGTGGT							
7520	7530	7540	7550	7560	7570	7580	7590

	1400	1410	1420	1430	1440	1450	1460
AATAACAACAATGGGTCAGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTA							
AATAACAACAATGGGTCAGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTA							
	7600	7610	7620	7630	7640	7650	7660

	1470	1480	1490	1500	1510	1520	1530
TATAAATATAAGTAGTAAATAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAGAGAGAGTGGTGCAG							
TATAAATATAAGTAGTAAATAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAGAGAGAGTGGTGCAG							
	7670	7680	7690	7700	7710	7720	7730

	1540	1550	1560	1570	1580	1590	1600
AGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGC							
AGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGC							
	7740	7750	7760	7770	7780	7790	7800

1610	1620	1630	1640	1650	1660	1670
-GCACGGTCAATGACGCTGACGCTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAAATTT						
TGCAC-GTCAATGACGCTGACGCTACAGGCCAGACAATTATTGTCTGATATAGTGCAGCAGCAGAACAAATTT						
7810	7820	7830	7840	7850	7860	7870

1680	1690	1700	1710	1720	1730	1740	1750
GCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAG							
GCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAG							
7880	7890	7900	7910	7920	7930	7940	7950

	1760	1770	1780	1790	1800	1810	1820
AATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGAAAACATCAT							
AATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGAAAACATCAT							
	7960	7970	7980	7990	8000	8010	8020

	1830	1840	1850	1860	1870	1880	1890
TTGCACCACTGCTGTGCTTGGAAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATAACATGAC							
TTGCACCACTGCTGTGCTTGGAAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATAACATGAC							
	8030	8040	8050	8060	8070	8080	8090

	1900	1910	1920	1930	1940	1950	1960
CTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACATTCTTAATTGAAGAATCGCAAAA							
CTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACATTCTTAATTGAAGAATCGCAAAA							
	8100	8110	8120	8130	8140	8150	8160

```

1970      1980      1990      2000      2010      2020      2030
CCAGCAGGAAAGGAATGAACAAGAAATTATTGGAAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAAACAT
|||||
CCAGCAGGAAAGGAATGAACAAGAAATTATTGGAAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAAACAT
8170      8180      8190      8200      8210      8220      8230

2040      2050      2060      2070      2080      2090      2100      2110
AACAAATTGGCTTGTGGTATATAAAAAATATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAAGAATAGTTTT
|||||
AACAAATTGGCTTGTGGTATATAAAAAATATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAAGAATAGTTTT
8240      8250      8260      8270      8280      8290      8300      8310

      2120      2130      2140      2150      2160      2170      2180
TGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCCAAC
|||||
TGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCCAAT
8320      8330      8340      8350      8360      8370      8380

      2190      2200      2210      2220      2230      2240      2250
CCCGAGGGGACCCGACAGGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCCG
|||||
CCCGAGGGGACCCGACAGGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCCG
8390      8400      8410      8420      8430      8440      8450

      2260      2270      2280      2290      2300      2310      2320
ATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGCCTTGTGCCTCTTCAGCTACCACCGCT
|||||
ATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGCC-TGTGCCTCTTCAGCTACCACCGCT
8460      8470      8480      8490      8500      8510      8520

2330      2340      2350      2360      2370      2380      2390
TGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACCTTCTGGGACGCGAGGGGGTGGGAAGCCCTCAAAT
|||||
TGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACCTTCTGGGACGCGAGGGGGTGGGAAGCCCTCAAAT
8530      8540      8550      8560      8570      8580      8590

2400      2410      2420      2430      2440
ATTGGTGGAAATCTCCTACAGTATTGGAGTCAGGAACTAAAG
|||||
ATTGGTGGAAATCTCCTACAGTATTGGAGTCAGGAACTAAAG
8600      8610      8620      8630      X

```

8. KUNZ-158-CL33, SEQ

HIVHXB2CG Human immunodeficiency virus type 1 (HXB2), comple

```

LOCUS      HIVHXB2CG      9718 bp ss-RNA      VRL      25-SEP-1987
DEFINITION Human immunodeficiency virus type 1 (HXB2), complete genome;
            HIV1/HTLV-III/LAV reference genome.
ACCESSION  K03455
KEYWORDS   TAR protein; acquired immune deficiency syndrome; complete genome;
            env gene; gag gene; long terminal repeat; pol gene; polyprotein;
            provirus; reverse transcriptase; trans-activator.
SOURCE      HTLV-III/LAV (isolate HXB2) proviral DNA.
ORGANISM    Human immunodeficiency virus type 1
            Viridae; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.
REFERENCE   1 (bases 1 to 653)
AUTHORS     Starcich,B., Ratner,L., Josephs,S.F., Okamoto,T., Gallo,R.C. and
            Wong-Staal,F.
TITLE       Characterization of long terminal repeat sequences of HTLV-III
JOURNAL      Science 227, 538-540 (1985)
STANDARD    full staff_review
REFERENCE   2 (bases 493 to 674; 9577 to 9718)
AUTHORS     Ratner,L., Haseltine,W., Patarca,R., Livak,K.J., Starcich,B.,
            Josephs,S.J., Doren,E.R., Rafalski,J.A., Whitehorn,E.A.,
            Baumalter,K., Ivanoff,L., Petteway,S.R. Jr., Pearson,M.L.,

```

Lautenberger, J. A. , Papas, T. S. , Ghrayeb, J. , Chang, N. T. , Gallo, R. C. and Wong-Staal, F.
 TITLE Complete nucleotide sequence of the AIDS virus, HTLV-III
 JOURNAL Nature 313, 277-284 (1985)
 STANDARD full staff_review
 REFERENCE 3 (bases 9761 to 9060)
 AUTHORS Fisher, A. G. , Ratner, L. , Mitsuya, H. , Marselle, L. M. , Harper, M. E. , Broder, S. , Gallo, R. C. and Wong-Staal, F.

TITLE Infectious mutants of HTLV-III with changes in the 3' region and markedly reduced cytopathic effects
 JOURNAL Science 233, 655-659 (1986)
 STANDARD full staff_review
 REFERENCE 4 (bases 1 to 9635)
 AUTHORS Ratner, L. , Fisher, A. , Jagodzinski, L. L. , Mitsuya, H. , Liou, R. -S. , Gallo, R. C. and Wong-Staal, F.

TITLE Complete nucleotide sequences of functional clones of the AIDS virus
 JOURNAL AIDS Res. Hum. Retroviruses 3, 57-69 (1987)
 STANDARD full staff_review
 REFERENCE 5 (sites; tat mRNA and other transcript boundaries)
 AUTHORS Arya, S. K. , Guo, C. , Josephs, S. F. and Wong-Staal, F.

TITLE Trans-activator gene of human T-lymphotropic virus type III (HTLV-III)
 JOURNAL Science 229, 69-73 (1985)
 STANDARD full staff_review
 REFERENCE 6 (sites; tat mRNA)
 AUTHORS Sodroski, J. , Patarca, R. , Rosen, C. , Wong-Staal, F. and Haseltine, W.

TITLE Location of the trans-activating region on the genome of human T-cell lymphotropic virus type III
 JOURNAL Science 229, 74-77 (1985)
 STANDARD full staff_review
 REFERENCE 7 (sites; mRNA splice sites)
 AUTHORS Rabson, A. B. , Daugherty, D. F. , Venkatesan, S. , Boulukos, K. E. , Berra, S. I. , Folks, T. M. , Feorino, P. and Martin, M.

TITLE Transcription of novel open reading frames of AIDS retrovirus during infection of lymphocytes
 JOURNAL Science 229, 1388-1390 (1985)
 STANDARD full staff_review
 REFERENCE 8 (sites; 27K antigen cds)
 AUTHORS Allan, J. S. , Coligan, J. E. , Lee, T. -H. , McLane, M. F. , Kanki, P. J. , Groopman, J. E. and Essex, M.

TITLE A new HTLV-III/LAV encoded antigen detected by antibodies from AIDS patients
 JOURNAL Science 230, 810-813 (1985)
 STANDARD full staff_review
 REFERENCE 9 (sites; gp160 and gp120 coding sequences)
 AUTHORS Allan, J. S. , Coligan, J. E. , Barin, F. , McLane, M. F. , Sodroski, J. G. , Rosen, C. A. , Haseltine, W. A. , Lee, T. H. and Essex, M.

TITLE Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III
 JOURNAL Science 228, 1091-1094 (1985)
 STANDARD full staff_review
 REFERENCE 10 (sites; regulatory sequences in the LTR)
 AUTHORS Rosen, C. A. , Sodroski, J. G. and Haseltine, W. A.

TITLE The location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat
 JOURNAL Cell 41, 813-823 (1985)
 STANDARD full staff_review
 REFERENCE 11 (review; bases 1 to 9718)
 AUTHORS Van Beveren, C. , Coffin, J. and Hughes, S.

TITLE Appendix B: HTLV-3 genome
 JOURNAL (in) Weiss, R. , Teich, N. , Varmus, H. and Coffin, J. (Eds.); RNA Tumor Viruses, Second Edition, 2: 1102-1123 and 1147-1148; Cold Spring Harbor Laboratory, New York (1985)
 STANDARD full staff_review
 REFERENCE 12 (sites; trans-activator function and TAR sequence)

AUTHORS Rosen, C. A. , Sodroski, J. G. , Goh, W. C. , Dayton, A. I. , Lippke, J. and Haseltine, W. A.

TITLE Post-transcriptional regulation accounts for the trans-activation of the human T-lymphotropic virus type III

JOURNAL Nature 319, 555-558 (1986)

STANDARD full staff_review

REFERENCE 13 (sites; pol coding sequence)

AUTHORS di Marzo Veronese, F. , Copeland, T. D. , DeVico, A. L. , Rahman, R. , Droszlan, S. , Gallo, R. C. and Sarngadharan, M. G.

TITLE Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV

JOURNAL Science 231, 1289-1291 (1986)

STANDARD full staff_review

REFERENCE 14 (sites; 23K sor gene product)

AUTHORS Kan, N. C. , Franchini, G. , Wong-Staal, F. , DuBois, G. C. , Robey, W. G. , Lautenberger, J. A. and Papas, T. S.

TITLE Identification of HTLV-III/LAV sor gene product and detection of antibodies in human sera

JOURNAL Science 231, 1553-1555 (1986)

STANDARD full staff_review

REFERENCE 15 (sites; pol NH2-terminal region)

AUTHORS Kramer, R. A. , Schaber, M. D. , Skalka, A. M. , Ganguly, K. , Wong-Staal, F. and Reddy, E. P.

TITLE HTLV-III gag protein is processed in yeast cells by the virus pol-protease

JOURNAL Science 231, 1580-1584 (1986)

STANDARD full staff_review

REFERENCE 16 (sites; sor 23K protein)

AUTHORS Lee, T. -H. , Coligan, J. E. , Allan, J. S. , McLane, M. F. , Groopman, J. E. and Essex, M.

TITLE A new HTLV-III/LAV protein encoded by a gene found in cytopathic retroviruses

JOURNAL Science 231, 1546-1549 (1986)

STANDARD full staff_review

REFERENCE 17 (sites; sor 23K protein)

AUTHORS Sodroski, J. , Goh, W. C. , Rosen, C. , Tartar, A. , Portetelle, D. , Bunny, A. and Haseltine, W.

TITLE Replicative and cytopathic potential of HTLV-III/LAV with sor gene deletions

JOURNAL Science 231, 1549-1553 (1986)

STANDARD full staff_review

REFERENCE 18 (sites; Sp1 binding sites in the promoter region)

AUTHORS Jones, K. A. , Kadonaga, J. T. , Luciw, P. A. and Tjian, R.

TITLE Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1

JOURNAL Science 232, 755-759 (1986)

STANDARD full staff_review

REFERENCE 19 (sites; acceptor and donor splice sites for tat and 27K)

AUTHORS Arya, S. K. and Gallo, R. C.

TITLE Three novel genes of human T-lymphotropic virus type III: Immune reactivity of their products with sera from acquired immune deficiency syndrome patients

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 83, 2209-2213 (1986)

STANDARD full staff_review

REFERENCE 20 (sites; deletion mutants in the tat gene)

AUTHORS Dayton, A. I. , Sodroski, J. G. , Rosen, C. A. , Goh, W. C. and Haseltine, W. A.

TITLE The trans-activator gene of the human T cell lymphotropic virus type III is required for replication

JOURNAL Cell 44, 941-947 (1986)

STANDARD full staff_review

REFERENCE 21 (sites; env gene conserved/variable regions; separate entries)

AUTHORS Willey, R. , Rutledge, R. A. , Dias, S. , Folks, T. , Theodore, T. , Buckler, C. E. and Martin, M. A.

TITLE Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome virus

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 83, 5038-5042 (1986)

STANDARD full staff_review
REFERENCE 22 (sites; trs cds boundaries)
AUTHORS Sodroski, J. , Goh, W. C. , Rosen, C. , Dayton, A. I. , Terwilliger, E. and Haseltine, W.
TITLE A second post-transcriptional trans-activator gene required for HTLV-III replication
JOURNAL Nature 321, 412-417 (1986)
STANDARD full staff_review
REFERENCE 23 (sites; trs cds boundaries)
AUTHORS Feinberg, M. B. ; Jarret, R. F. ; Aldovini, A. , Gallo, R. C. and Wong-Staal, F.
TITLE HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA
JOURNAL Cell 46, 807-817 (1986)
STANDARD full staff_review
REFERENCE 24 (sites; env gene conserved/variable regions; separate entries)
AUTHORS Starcich, B. R. , Hahn, B. H. , Shaw, G. M. , McNeely, P. D. , Modrow, S. , Wolf, H. , Parks, E. S. , Parks, W. P. , Josephs, S. F. , Gallo, R. C. and Wong-Staal, F.
TITLE Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS
JOURNAL Cell 45, 637-648 (1986)
STANDARD full staff_review
REFERENCE 25 (sites; tar or transactivator target)
AUTHORS Wright, C. M. ; Felber, B. K. , Paskalis, H. and Pavlakis, G. N.
TITLE Expression and characterization of the trans-activator of HTLV-III/LAV virus
JOURNAL Science 234, 988-992 (1986)
STANDARD full staff_review
REFERENCE 26 (sites; 3' orf mutations)
AUTHORS Terwilliger, E. , Sodroski, J. G. , Rosen, C. A. and Haseltine, W. A.
TITLE Effects of mutations within the 3' orf open reading frame region of human T-cell lymphotropic virus type III (HTLV-III/LAV) on replication and cytopathogenicity
JOURNAL J. Virol. 60, 754-760 (1986)
STANDARD full staff_review
REFERENCE 27 (sites; pol p34 terminus)
AUTHORS Lightfoote, M. M. , Coligan, J. E. , Folks, T. M. , Fauci, A. S. , Martin, M. A. and Venkatesan, S.
TITLE Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus
JOURNAL J. Virol. 60, 771-775 (1986)
STANDARD full staff_review
REFERENCE 28 (sites; promoter, TAR, tat-III mutants)
AUTHORS Muesing, M. A. , Smith, D. H. and Capon, D. J.
TITLE Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein
JOURNAL Cell 48, 691-701 (1987)
STANDARD full staff_review
REFERENCE 29 (sites; envelope protein epitopes)
AUTHORS Modrow, S. , Hahn, B. H. , Shaw, G. M. , Gallo, R. C. , Wong-Staal, F. and Wolf, H.
TITLE Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: Prediction of antigenic epitopes in conserved and variable regions
JOURNAL J. Virol. 61, 570-578 (1987)
STANDARD full staff_review
REFERENCE 30 (sites; trs/art protein)
AUTHORS Goh, W. C. , Sodroski, J. G. , Rosen, C. A. and Haseltine, W. A.
TITLE Expression of the art gene protein of human T-lymphotropic virus type III (HTLV-III/LAV) in bacteria
JOURNAL J. Virol. 61, 633-637 (1987)
STANDARD full staff_review
REFERENCE 31 (sites; inducible enhancer element)

AUTHORS Nabel, G. and Baltimore, D.
TITLE An inducible transcription factor activates expression of human immunodeficiency virus in T cells
JOURNAL Nature 326, 711-713 (1987)
STANDARD full staff_review
REFERENCE 32 (bases 5611 to 5611; revises [4])
AUTHORS Ratner, L.
JOURNAL Unpublished (1987) Washington U Med School, St. Louis, MO
STANDARD full staff_review
REFERENCE 33 (sites; long terminal repeat)
AUTHORS Paterca, R., Heath, C., Goldenberg, G. J., Rosen, C. A., Sodroski, J. G., Haseltine, W. A. and Hansen, U. M.
TITLE Transcription directed by the HIV long terminal repeat in vitro
JOURNAL AIDS Res. Hum. Retroviruses 3, 41-55 (1987)
STANDARD full staff_review
REFERENCE 34 (sites; R orf)
AUTHORS Wong-Staal, F., Chanda, P. K. and Ghrayeb, J.
TITLE Human immunodeficiency virus: the eighth gene
JOURNAL AIDS Res. Hum. Retroviruses 3, 33-39 (1987)
STANDARD full staff_review
REFERENCE 35 (sites; sor)
AUTHORS Fisher, A. G., Ensoli, B., Ivanoff, L., Chamberlain, M., Petteway, S., Ratner, L., Gallo, R. C. and Wong-Staal, F.
TITLE The sor gene of hiv-1 is required for efficient virus transmission in vitro
JOURNAL Science 237, 888-893 (1987)
STANDARD full staff_review
COMMENT Sequence for [3] kindly provided in computer-readable form by L. Ratner, 19-AUG-1986.
The HXB2 sequence is being used as a reference genome for all the HIV entries because it has been derived from a demonstrably infectious clone. Hence not all of the "sites" references above were concerned with this isolate.

FEATURES	from	to/span	description
pept	789	2291	gag polyprotein
pept	/ 2357	5085	pol polyprotein (NH2-terminus uncertain; AA at 2357)
pept	5040	5618	sor 23K protein
pept	5558	5794	R (ORF) protein
pept	5830	6044	tat protein, exon 2 (first expressed exon)
	5378	6423	tat protein, exon 3
pept	5558	6044	trs protein, exon 2 (first expressed exon)
	5378	6652	trs protein, exon 3
pept	6224	8784	envelope polyprotein
pept	5786	5167	27K protein (premature termination)
mRNA	455	8635	HXB2 genomic mRNA
pre-msg	455	8635	tat, trs, 27K subgenomic mRNA
IVS	6045	8377	tat intron 1
IVS	6045	8377	trs intron 2
IVS	6045	8377	27K mRNA intron 2
IVS	743	5776	tat, trs, 27K mRNA intron 1
IVS	6045	8377	tat, trs intron 2
LTR	1	634	5' LTR
LTR	9085	9718	3' LTR
rpt	454	551	R repeat 5' copy
rpt	9538	8635	R repeat 3' copy
binding	377	386	Sp1 binding site III
binding	388	387	Sp1 binding site II
binding	398	408	Sp1 binding site I
binding	638	633	primer (Lys-tRNA) binding site
revision	5611	5611	g in [32]; a in [4]
signal	8611	8616	HXB2 mRNA polyadenylation signal
BASE COUNT	3411 a	1773 c	2370 g 2164 t
ORIGIN	435 bp upstream of PvuII site; 5' end of proviral genome.		

Initial Score = -1258 Optimized Score = -2164 Significance = 0.00

Residue Identity	=	89%	Matches	=	2231	Mismatches	=	162
Gaps	=	100	Conservative Substitutions	=			=	0

X

10 20 30 40 50 60
ATGAGAGTGA-AGEAGAA-ATATCAGCACTTTGTGGAGA-TGGGGGTGGAAATGGGGCAC-CATGCTCCTTGG
| | | | | | | | | | | | | | | | | |
CTAATAGAYAGAGCAGAGACAGTGGCAAT-----GAGAGTGAAGGAGAAATATCAGCACTTTGTGGAGATGGG
6200 6210 6220 6230 6240 6250 6260

70 80 90 100 110 120
GATATTGATG-----AT-CT-GTAGTGCTACAGAAAAATTGT-GGGTCACAG-TCTATTATGGGGTAC--
| | | | | | | | | | | | | | | | | |
GGTGGAGATGGAGACCATECTCCCTTEGGATGTTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTGCACAG
6270 6280 6290 6300 6310 6320 6330

130 140 150 160 170 180
-CT---GTGTGGAA-----GGAGSCAA-CCACCA-CTCTATTTTGTGCATCAGATGCTAAAGCATATGAT
| | | | | | | | | | | | | | | | | |
TCTAITATGEGGTACCGTGTGTGGAGGAGGCAACCACCACTCTATTTTG-----TGC---ATCAGATGCT
6340 6350 6360 6370 6380 6390

190 200 210 220 230 240
A---CAGAGG-TACATA-----AT--GTTTGGGCCACACATGCCTG--TGTACCCACAGA-CCCCAACCCAC
| | | | | | | | | | | | | | | | | |
AAAGCATATGATACAGAGGTACATAATGTTTTGGGC-CACA--CATGCCTGTGTACCCACAGACCCCAACC-C
6400 6410 6420 6430 6440 6450 6460

250 260 270 280 290 300 310
AAGAGTAGTATTGCTAAATGTGACAGAAAAATTTTAACATGTGGAAAA--ATGACATGGTAGAACAGATG-C
| | | | | | | | | | | | | | | | | |
A-CAGGAAGTA--GTA-TTG-----GTAAT-GTGACA----GAAAATTTTGACAT-GTGGAA-AAATGAC
6470 6480 6490 6500 6510 6520

320 330 340 350 360 370 380
ATGAGGATATAATCAG-TTTATG-GGATCAAAGCCTAAAGCCATGTG-TAAAA--TTAACCCCCTCTGTGT
| | | | | | | | | | | | | | | | | |
AT---GGTAGAA-CAGATGCATGAGGATATAA---TCAGTTTATGGGATCAAAGCCTAAAGCCA---TGTGT
6530 6540 6550 6560 6570 6580

390 400 410 420 430 440
TAGTTTAA-AGTGCACTGATTTGG--GGAAATGCTACTAAT--ACCAATACTAGTAATACCAATAGTAGTA
| | | | | | | | | | | | | | | | | |
AAAATTAACCCCCTCTGTGTTAGTTTAAAGTGC-ACTGATTTGAAGAATGATACTAATACCAATAGTAGTA
6590 6600 6610 6620 6630 6640 6650

450 460 470 480 490 500 510
GCGGGGAAATGATGATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAAGNATAAGAGGTA
| | | | | | | | | | | | | | | | | |
GCGGGAGAAATGATAATGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAAGCATAAGAGGTA
6660 6670 6680 6690 6700 6710 6720

520 530 540 550 560 570 580
AGGTGCAGAAAGAAATATGCATTTT TTTATAAACTTGATATAATACCAATAGATAATGATACTACCAGCTATA
| | | | | | | | | | | | | | | | | |
AGGTGCAGAAAGAAATATGCATTTT TTTATAAACTTGATATAATACCAATAGATAATGATACTACCAGCTATA
6730 6740 6750 6760 6770 6780 6790

590 600 610 620 630 640 650 660
CGTTGACAAATGTACACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATAC
| | | | | | | | | | | | | | | | | |
GCTTGACAAATGTACACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATAC
6800 6810 6820 6830 6840 6850 6860

670 680 690 700 710 720 730
ATTTATTGTGCCCGGGCTGGT TTTGGATTCTAAATGTAATAATAAGACGTTCAATGGAAACAGGACCATGTA
| | | | | | | | | | | | | | | | | |

ATTATTGTGCCCCGGCTGGT TTTCCGATTCTAAAATGTAATAAAGACGTTCAATGGAACAGGCATGTA
6870 6880 6890 6900 6910 6920 6930 6940

740 750 760 770 780 790 800
CAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCA
|||||
CAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCA
6950 6960 6970 6980 6990 7000 7010

810 820 830 840 850 860 870
GTCTAGCABAAGAAAGTAGTAATTAGATCTGCCAATTTACAGACAATGCTAAAACCATAATAGTACAGC
|||||
GTCTAGCABAAGAAAGTAGTAATTAGATCTGTCAATTTACGGACAATGCTAAAACCATAATAGTACAGC
7020 7030 7040 7050 7060 7070 7080

880 890 900 910 920 930 940
TGAACCAATCTGTAGAAATTAATTGTACAAGACCCACAACAATACAAGAAAAAGTATCCGTATCCAGAGGG
|||||
TGAACACATCTGTAGAAATTAATTGTACAAGACCCACAACAATACAAGAAAAAGTATCCGTATCCAGAGAG
7090 7100 7110 7120 7130 7140 7150

950 960 970 980 990 1000 1010 1020
GACCAGGGAGAGCATTGTTTACAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAG
|||||
GACCAGGGAGAGCATTGTTTACAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAG
7160 7170 7180 7190 7200 7210 7220

1030 1040 1050 1060 1070 1080 1090
CAAAATGCAATGCCACTTTAAACACAGATAGCTAGCAAATTAAGAGAACAATTTGGAAATAATAAAACAATAA
|||||
CAAAATGGAATAACACTTTAAACACAGATAGCTAGCAAATTAAGAGAACAATTCGGAAATAATAAAACAATAA
7230 7240 7250 7260 7270 7280 7290 7300

1100 1110 1120 1130 1140 1150 1160
TCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTCT
|||||
TCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTCT
7310 7320 7330 7340 7350 7360 7370

1170 1180 1190 1200 1210 1220 1230
ACTGTAATTCACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACA
|||||
ACTGTAATTCACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACA
7380 7390 7400 7410 7420 7430 7440

1240 1250 1260 1270 1280 1290 1300
CTGAGAGGAGTGCACAATCACACTCCCATGCAGAAATAAAACAATTTATAACATGTGGCAGGAAGTAGGAA
|||||
CTGAGAGGAGTGCACAATCACCTCCCATGCAGAAATAAAACAATTTATAACATGTGGCAGAAAGTAGGAA
7450 7460 7470 7480 7490 7500 7510

1310 1320 1330 1340 1350 1360 1370 1380
AAGCAATGTATGCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAAATATTACAGGGCTGCTATTAACAA
|||||
AAGCAATGTATGCCCTCCCATCAGTGGACAAATTAGATGTTTCATCAAATATTACAGGGCTGCTATTAACAA
7520 7530 7540 7550 7560 7570 7580

1390 1400 1410 1420 1430 1440 1450
GAGATGGTGTATATAACAAATGAGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGA
|||||
GAGATGGTGTATATAACAAATGAGTCCGAGATCTTCAGACTTGGAGGAGGAGATATGAGGGACAATTGGA
7590 7600 7610 7620 7630 7640 7650 7660

1460 1470 1480 1490 1500 1510 1520
GAGTGAATATATTAATATTAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAGAGAA
|||||

GAAGTGAATTAATATAATATTAAGTAGTAAGAAATTGAACCATAGGAGTAGCACCCACCAAGGCCAAGAGAA
7670 7680 7690 7700 7710 7720 7730

1530 1540 1550 1560 1570 1580 1590
GASTGGTGCAGAGASANAANAAGAGCAGTGGGAATAGGAGCTTTGTTCCCTTGGGTTCTTGGGAGCAGCAGGAA
.....
GASTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCCTTGGGTTCTTGGGAGCAGCAGGAA
7740 7750 7760 7770 7780 7790 7800

1600 1610 1620 1630 1640 1650 1660
GCACATATGGGCGCAGCTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGC
.....
GCATATGGGCGCAGCTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGC
7810 7820 7830 7840 7850 7860 7870

1670 1680 1690 1700 1710 1720 1730 1740
AGAAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGC
.....
AGAAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGC
7880 7890 7900 7910 7920 7930 7940

1750 1760 1770 1780 1790 1800 1810
TCCAGGCAGAGAACTCTGGCTGTGGAAGATACCTAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTG
.....
TCCAGGCAGAGAACTCTAGCTGTGGAAGATACCTAAGGATCAACAGCTCCTAGGGATTTGGGGTTGCTCTG
7950 7960 7970 7980 7990 8000 8010 8020

1820 1830 1840 1850 1860 1870 1880
GAAAACATCATTTGCACCACTGCTGTGCCCTTGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGA
.....
GAAAACATCATTTGCACCACTGCTGTGCCCTTGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATCTGGA
8030 8040 8050 8060 8070 8080 8090

1890 1900 1910 1920 1930 1940 1950
ATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACATTCTTAATTGAAG
.....
ATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCTTAATTGAAG
8100 8110 8120 8130 8140 8150 8160

1960 1970 1980 1990 2000 2010 2020
AATCGCAAAACCGAGCAGAGAAAGAAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATT
.....
AATCGCAAAACCGAGCAGAGAAAGAAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATT
8170 8180 8190 8200 8210 8220 8230

2030 2040 2050 2060 2070 2080 2090 2100
GGTTTAACATACAAATTGGCTGTGGTATATAAAATATTTCATAATGATAGTAGGAGGCTTGGTAGGTTTAA
.....
GGTTTAACATACAAATTGGCTGTGGTATATAAAATATTTCATAATGATAGTAGGAGGCTTGGTAGGTTTAA
8240 8250 8260 8270 8280 8290 8300

2110 2120 2130 2140 2150 2160 2170
GAATAGTTTCTTCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCC
.....
GAATAGTTTCTTCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCC
8310 8320 8330 8340 8350 8360 8370 8380

2180 2190 2200 2210 2220 2230 2240
ACCTCCCACACCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACA
.....
ACCTCCCACATCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACA
8390 8400 8410 8420 8430 8440 8450

2250 2260 2270 2280 2290 2300 2310
GATCCATTGATTAATGAACGATCCTTAGCACTTATCTGGGACGATCTGCGGAGCCTTGTGCCTCTTCAGC
.....

GATCGATTGGAT TAGTGAACGATCCTTGCCACTTATCTGGGACGATCTGCGGAGCC-TGTGCCTCTTCAGC
8460 8470 8480 8490 8500 8510 8520

2320 2330 2340 2350 2360 2370 2380
TACGACCGCTTGGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGACGCGAGGGGGTGGGAA
|||||
TACGACCGCTTGGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGACGCGAGGGGGTGGGAA
8530 8540 8550 8560 8570 8580 8590

2390 2400 2410 2420 2430 2440
GCCCTCAAATAT TGGTGGAACTCTCCTACAGTATTGGAGTCAGGAACTAAAG
|||||
GCCCTCAAATAT TGGTGGAACTCTCCTACAGTATTGGAGTCAGGAACTAAAG
8600 8610 8620 8630 8640 X

9. KUNZ-158-CL33.529

HIVSC Human immunodeficiency virus type 1, isolate SC (3

LOCUS HIVSC 4273 bp ss-RNA VRL 15-JUN-1989
DEFINITION Human immunodeficiency virus type 1, isolate SC (3' end of genome).
ACCESSION M17450
KEYWORDS .
SOURCE Human immunodeficiency virus type 1 (HIV-1), isolate SC, provial
DNA
ORGANISM Human immunodeficiency virus type 1
Virus: ss-RNA enveloped viruses; Retroviridae;
Lentivirinae.
REFERENCE 1 (bases 1 to 4273)
AUTHORS Gurgo,C., Guo,H.-G., Franchini,G., Aldovini,A., Collalti,E.,
Farrell,K., Wong-Staal,F., Gallo,R.C. and Reitz,M.S. Jr.
TITLE Envelope sequences of two new United States HIV-1 isolates
JOURNAL Virology 164, 531-536 (1988)
STANDARD full staff_review
COMMENT kindly made available in computer readable form by Marv Reitz,
N.C.I., Bethesda, MD 20892 U.S.A. This isolate was taken from a
California AIDS patient in 1984. There is an in-frame stop codon at
position 3212 of the envelope coding sequence; the nef cds is
uncertain beyond position 4049. A stop codon, 'taa,' in-frame with
the nef sequence does exist at positions 4224-4226.

FEATURES	from	to/span	description
pept	(1	330	vif protein (partial; AA at 1)
pept	270	560	vpr protein
pept	541	755	tat protein, exon 2 (first expressed exon)
	3088	3178	tat protein, exon 3 (AA at 3090)
pept	680	755	rev protein, exon 2 (first expressed exon)
	3089	3353	rev protein, exon 3 (AA at 3091)
pept	772	864	vpu protein (premature termination)
pept, ps	935	3305	envelope polyprotein (premature stop at 3212)
pept	3507	4226	nef protein
pre-msg (1	4273	genomic mRNA
pre-msg (1	4273	tat, rev, nef subgenomic mRNA
IVS	(1	437	tat, rev, nef subgenomic mRNA intron 1
IVS	756	3088	tat cds intron 2
IVS	756	3088	rev cds intron 2
IVS	756	3088	tat, rev, nef subgenomic mRNA intron 2
LTR	3795	4273	3' LTR
rpt	4248	4273	R repeat 3' copy
site	3212	3214	premature stop (tag) in env cds
BASE COUNT	1447 a	760 c	1053 g 1013 t
ORIGIN			

Initial Score = 1159 Optimized Score = 2139 Significance = 0.00
Residue Identity = 87% Matches = 2194 Mismatches = 203
Gaps = 97 Conservative Substitutions = 0

[illegible]

770 780 790 800 810 820 830
 TAGGCCAGTATGATCAACTCAACTGCTGTTGAA TGGCAGTCTAGCAGAGAAGAGGTTAGTAATTAGATCTGC
 |||||
 TAGGCCAGTATGATCAACTCAACTGCTGTTGAA TGGCAGTCTAGCAGAGAAGAGGTTAGTACTTAGGTCTGA
 1690 1700 1710 1720 1730 1740 1750

840 850 860 870 880 890 900 910
 GAATTTTCACAGGCAATGCTAAAACCATTAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACC
 |||||
 GAATTTTCACAGGCAATGCTAAAACCATTAATAGTACAGCTGAAGAGAGCTGTAGAAATTAATTGTACAAGGCC
 1760 1770 1780 1790 1800 1810 1820 1830

920 930 940 950 960 970 980
 CAACACCAATACAGAGGAAATATCCGTATCCAGAGGGGACCAGGGAGAGCATTGTTACAATAGGA---AA
 |||||
 CAACACCAATACAGAGGAAATATACATAT---AG---GACCAGGGAGAGCATTGTTATGCAACAGGAGACAT
 1840 1850 1860 1870 1880 1890

990 1000 1010 1020 1030 1040 1050
 AATAGGAAATATGACACAGGCAATGTTAATAGTAGAGCAAAATGCAATGCCACTTTAAACAGATAGC
 |||||
 AATAGGAGATATGACACAGGCAATGTTAATAGTAGAGCAAAATGGAATAACACTTTAAACAGATAGT
 1900 1910 1920 1930 1940 1950 1960

1060 1070 1080 1090 1100 1110 1120
 TACCAAAATTAAGAGAAACAATTGGGAAATAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGA
 |||||
 TATAAAATTAAGAGACCAATTTG---AGAATAAAACAATAATCTTTAATCGATCCTCAGGAGGAGACCCAGA
 1970 1980 1990 2000 2010 2020 2030

1130 1140 1150 1160 1170 1180 1190
 AATTGTAAAGCAGCAATTTTAAATGTTGGAGGGGAATTTTCTACTGTAATTCAACACAACCTGTTTAAATAGTAC
 |||||
 AATTGTAAAGCAGCAATTTTAAATGTTGGAGGGGAATTTTCTACTGTAATTCAACACAACCTGTTTAAATAGTAC
 2040 2050 2060 2070 2080 2090 2100

1200 1210 1220 1230 1240
 TTGTTTAAATAGTACT---TGG---AGTA---CTGAAGG---GTCA-AATAACAC---TGAAG---GAAG
 |||||
 TTGG---AATGTTACTGAAAGGTCAAAATAACACTGGAGGAAATGACACAATCACCCCTCCCATGCAGAAATAA
 2110 2120 2130 2140 2150 2160 2170

1250 1260 1270 1280 1290 1300 1310
 TGACACAAT-CACA-CTCCCATGAGAAATAAAACAATTTATAAACATGTGGCAGGAAGTAGGAAAAGCAATG
 |||||
 AGAAATTATAAATCATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATCAA---AGGACAAG---TTA
 2180 2190 2200 2210 2220 2230 2240

1320 1330 1340 1350 1360 1370 1380
 TATCCCCCTCCCATGAGGCAAAATAGATGTTTCATCAAAATATTACAGGGCTGCTATTAACAAGAGATGGT
 |||||
 AATGTTCAACAAAT---ATTACA---GGGCTGCT-AT-TAACAAGAGATGG-TGGTAATAGCAAGA-ATGGT
 2250 2260 2270 2280 2290 2300

1390 1400 1410 1420 1430 1440 1450 1460
 GGTGATAACACAAATGGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAA
 |||||
 AGCAAGAATGAGAA---CACAGAGATCTTTAGACCGGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAA
 2310 2320 2330 2340 2350 2360 2370

1470 1480 1490 1500 1510 1520 1530
 TTATATAAATATAAAGTAGTAAATTTGAACCATTAGGAGTAGCACCCACCAAGGCAAGAGAGAGTGGTG
 |||||
 TTATATAAATATAAAGTAGTAAATTTGAACCATTAGGAGTAGCACCCACCAAGGCAAGAGAGAGTGGTG
 2380 2390 2400 2410 2420 2430 2440

[illegible]

```

3330      3340      3350      2360      2370      2380      2390
CORCTTGAGAGCACTTACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCCCT
|||||
CORCTTGAGAGCACTTACTCTTGATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCCCT
3250      3260      3270      3280      3290      3300      3310

2400      2410      2420      2430      2440
CAATATTGCTGGAATCTCCTACAGTATTEGAGTCAGGAACTAAAG
|||||
CAATATTGCTGGAATCTCCTACAGTATTEGAGTCAGGAACTAAAG
3320      3330      3340      3350      X

```

10. KUNZ-158-CL35.588

HIVMAL Human immunodeficiency virus type 1, isolate MAL,

LOCUS	HIVMAL	9229 bp ss-RNA	VR	15-JUN-1989
DEFINITION	Human immunodeficiency virus type 1, isolate MAL, complete genome.			
ACCESSION	U03456			
KEYWORDS	.			
SOURCE	Human immunodeficiency virus type 1 (HIV-1), isolate MAL, proviral DNA, clone M-H11.			
ORGANISM	Human immunodeficiency virus type 1 Virusae; ss-RNA enveloped viruses; Retroviridae; Lentivirinae.			
REFERENCE	1. (bases 1 to 9229)			
AUTHORS	Alizon,M., Vain-Hobson,S., Montagnier,L. and Sonigo,P.			
TITLE	Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isolates from African patients			
JOURNAL	Cell 46, 63-74 (1986)			
STANDARD	full staff_review			
COMMENT	A region similar to the vpu cds of HIVELI exists from positions 5536 (starts with 'ata' instead of 'atg') to 5881. To date it has not been possible to construct an infectious clone for the MAL viral genome (Keith Peden, Johns Hopkins University School of Medicine, Baltimore, MD).			
FEATURES	from	to/span	description	
pept	350	1857	gag polyprotein	
pept	(1663	4671	pol polyprotein (NH2-terminus uncertain; AA at 1663)	
pept	4616	5184	vif protein	
pept	5134	5424	vpr protein	
pept	5405	5613	tat protein, exon 2 (first expressed exon)	
	7959	8007	tat protein, exon 3 (AA at 7960)	
pept	5544	5619	rev protein, exon 2 (first expressed exon)	
	7959	8236	rev protein, exon 3 (AA at 7961)	
pept	(5536	5881	vpu protein (AA at 5636)	
pept	5789	6373	envelope polyprotein precursor	
pept	8450	9009	nef protein	
premsg	1	9228	genomic mRNA	
premsg	1	9228	tat, rev, nef subgenomic mRNA	
IVS	506	5351	tat, rev, nef subgenomic mRNA intron 1	
IVS	5620	7558	tat cds intron 2	
IVS	3520	7958	rev cds intron 2	
IVS	5620	7958	tat, rev, nef subgenomic mRNA intron 2	
LTR	(1	177	5' LTR	
LTR	8673) 9228	3' LTR	
rpt	(1	95	R repeat 5' copy	
rpt	9133	9228	R repeat 3' copy	
binding	179	196	primer (Lys-tRNA) binding site	
signal	9204	9203	mRNA polyadenylation signal	
BASE COUNT	3355 a	1627 c	2204 g	2043 t
ORIGIN	Cap site of genomic RNA.			

Initial Score = 916 Optimized Score = 2041 Significance = 0.00
Residue Identity = 84% Matches = 2066 Mismatches = 349

720 730 740 750 760 770 780
 TTCAATGGAGCAGGACCATGTACAGATGTCTAGCAGCAGTACAATGTACACATGGAATTAGGCCAGTAGTATCA
 ||||| | | | | | | | | | | | | | | | | | |
 TTCAATGGAGCAGGAATATGTAAAAATGTCTAGTACAGTACAATGTACACATGGAATTAGGCCAGTGGTGTCA

6510 6520 6530 6540 6550 6560 6570 6580
780 800 810 820 830 840 850
ACTCAACTGCTGTTGAATGGCAGTCTAGCAGAAGAGAGGTAGTAATTAGATCTGCCAATTTTCACAGACAAT
|||||
ACTCAACTGCTGTTGAATGGCAGTCTAGCAGAAGAGAGATAATGATTAGATCTGAAAATCTCACAGACAAT
6590 6600 6610 6620 6630 6640 6650

860 870 880 890 900 910 920
GCTAAAACCATAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGA
|||||
ACTAAAACATAATAGTACAGCTTAATGAACTGTAACAATTAATTGTACAAGGCCTGGAAACAATACAAGA
6660 6670 6680 6690 6700 6710 6720

930 940 950 960 970 980 990
AAAAGTATCCGATTCAGAGGGGACCAGGGAGAGCATTGTGTTACAATAGGAAAAATAGGAAATATGAGACAA
|||
AGAGGGATACATTTT-----GCCCCAGGGCAAGCACTCTATACAACAGGGATAGTAGGAGATATAAGAAGA
6730 6740 6750 6760 6770 6780 6790

1000 1010 1020 1030 1040 1050 1060 1070
GCACATTGTAACTATTAGTAGAGCAAAATGCAATGCCACTTTAAACAGATAGCTAGCAAATTAAGAGAACAA
|||||
GCATATTGTACTATTAATGAACAGAAATGGGATAAACTTTACAACAGGTAGCTGTAAACTAGGA-AGC--
6800 6810 6820 6830 6840 6850 6860

1080 1090 1100 1110 1120 1130 1140
TTTGGAAATAATGAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTT
|||
CTTCTTAACAAACAAAATAATTTTTAATTATCCTCAGGAGGGGACCCAGAAATTACAACACACAGTTTT
6870 6880 6890 6900 6910 6920 6930

1150 1160 1170 1180 1190 1200 1210
AATTGTGGAGGGGAATTTTCTACTGTAATTCAACACAACACTGTTTAATAGTACTTGGTTTAATAGTACTTGG
|||||
AATTGTAGAGGGGAATTTTTCTACTGTAATACATCAAACTGTTTAATAGTACATGGCAGAATAAT-GGTGC
6940 6950 6960 6970 6980 6990 7000

1220 1230 1240 1250 1260 1270 1280
AGTACTGAAGGGTCAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCAGAATAAAACAATTTATA
|||
AAGACT-AAG-----TAATAGCACAGAGTCAACTGGTAGTATCACACTCCCATGCAGAATAAAACAATTTATA
7010 7020 7030 7040 7050 7060 7070

1290 1300 1310 1320 1330 1340 1350
AACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAAT
|||
AATATGTGGCAGAAAACAGGAAAAGCTATGTATGCCCTCCCATCGCAGGAGTCATCAACTGTTTATCAAT
7080 7090 7100 7110 7120 7130 7140

1360 1370 1380 1390 1400 1410 1420
ATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAACA---ACAAT-GGGTC--CGAGATCTTCAGACCT
|||||
ATTACAGGGCTGCTATTAACAAGAGATGGTGGAAATAGTAGTGACAATAGTGACAATGAGACCTTAAGACCT
7150 7160 7170 7180 7190 7200 7210

1430 1440 1450 1460 1470 1480 1490
GGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTA
|||||
GGAGGAGGAGATATGAGGGACAATTGGATAAGTGAATTATATAAATATAAAGTAGTAAAGATTGAACCCCTA
7220 7230 7240 7250 7260 7270 7280

1500 1510 1520 1530 1540 1550 1560
GGAGTAGCACCCACCAAGGCCAAAGAGAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTG
|||||
GGAGTAGCACCCACCAAGGCCAAAGAGAGAGTGGTGGAAAGAGAAAAAGAGCAATAGGACTAGGAGCCATG
7290 7300 7310 7320 7330 7340 7350

[illegible]

8080 8090 8100 8110 8120 8130 8140

2360 2370 2380 2390 2400 2410 2420 2430

TGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCCCTCAAATATTGGTGGGAATCTCCTACAGTATTGGAGTCA

|||||

TGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCCCTCAAATATCTGTGGGAATCTCCTGCAATATTGGGGTCA

8150 8160 8170 8180 8190 8200 8210 8220

X

GGAACTAAAG

|||||

GGAACTGAGG

8230

Results file kunz-158-cl33umbl.res made by sheppard on Thu 8 Mar 90 11:32:28-PST.

Query sequence being compared: KUNZ-158-CL33.SEQ
Number of sequences searched: 3460
Number of scores above cutoff: 10

Results of the initial comparison of KUNZ-158-CL33.SEQ with:
Data bank : UEMBL 21, all entries

10000-
-
N -
U 5000-
M -
B -
E --
R -

O --
 F 1000% --
 S --
 E 500--
 Q -- *
 U --
 E --
 N --
 C --
 E --
 S 100--

50--

10--

5--

0--

SCORE	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
STDEV	0	2	1	7	5																

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	145		
Randomisation group	0		

Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	40	38	46.71

Times:	CPU	Total Elapsed
	00:07:23.04	00:32:23.00

Number of residues:	5125998
Number of sequences searched:	3460
Number of scores above cutoff:	10

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 33 standard deviations above mean ****						
1. REHTLV3	Human T-cell leukaemia type II	9748	1872	2176	39.22	0
2. HIVH306	Human T-cell lymphotropic viru	9749	1872	2176	39.22	0
**** 25 standard deviations above mean ****						
3. HIVELI0F	Human lymphadenopathy virus (E	9176	1246	1893	25.82	0
**** 18 standard deviations above mean ****						
4. HIVMAL06	Human lymphadenopathy virus (M	9229	916	2041	18.75	0
**** 5 standard deviations above mean ****						
5. HIV2R0DX	Human immunodeficiency virus t	9671	306	1185	5.69	0
6. REHTLV4G	STLV-3 (HTLV-4) partial provir	5391	299	1247	5.54	0
7. RESIVAXX	Simian immunodeficiency virus	9264	294	1240	5.44	0
8. RESIVMXX	Simian immunodeficiency virus	9646	290	1240	5.35	0
**** 4 standard deviations above mean ****						
9. RESIV261	Simian Immunodeficiency virus	1142	259	565	4.69	0
**** 3 standard deviations above mean ****						
10. M15127	Figure 1. Structure of the art	306	184	298	3.08	0

The scores below are sorted by optimized score.

Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. REHTLV3	Human T-cell leukaemia type II	9748	1872	2176	0.00	0
2. HIVH306	Human T-cell lymphotropic viru	9749	1872	2176	0.00	0
3. HIVMAL06	Human lymphadenopathy virus (M	9229	916	2041	0.00	0
4. HIVELI06	Human lymphadenopathy virus (E	9176	1246	1893	0.00	0
5. REHTLV4G	STLV-3 (HTLV-4) partial provir	5391	299	1247	0.00	0
6. RESIVMXX	Simian immunodeficiency virus	9646	290	1240	0.00	0
7. RESIVAXX	Simian immunodeficiency virus	9264	294	1240	0.00	0
8. HIV2R0DX	Human immunodeficiency virus t	9671	306	1185	0.00	0
9. RESIV261	Simian Immunodeficiency virus	1142	259	565	0.00	0
10. M15127	Figure 1. Structure of the art	306	184	298	0.00	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	2176	2177	0.00

Times:	CPU	Total Elapsed
	00:00:00.00	00:00:00.00

Number of residues:	73322
Number of sequences optimized:	10

1. KUNZ-158-CLASS. SER

REHTLV3 Human T-cell leukaemia type III (HTLV-III) provira

ID	REHTLV3	standard RNA: 9748 BP.
XX		
AC	XC1782	
XX		

DT 03-SEP-1987 (an correction)
 DT 01-SEP-1987 (an correction)
 DT 03-AUG-1987 (an correction)
 DT 29-OCT-1986 (minor modification)
 DT 05-NOV-1985 (NW added)
 DT 26-MAR-1985 (first entry)
 XX
 DE Human T-cell leukemia type III (HTLV-III) proviral genome
 DE (AIDS virus for acquired immune deficiency syndrome)
 XX
 KW acquired immune deficiency syndrome; direct repeat; endonuclease;
 KW glycoprotein; inverted repeat; protease; provirus;
 KW reverse transcriptase; terminal repeat.
 XX
 OS Human T-cell leukemia virus type III
 OC Viridae; ss-RNA enveloped viruses; Retroviridae.
 XX
 RN [1] (bases 1-8748)
 RA Ratner L., Haseltine W., Patarca R., Livak K. J., Starcich B. R.,
 RA Josephs S. F., Doran E. R., Rafalski J. A., Whitehorn E. A.,
 RA Baumeister K., Ivanoff L., Petteway S. R. Jr., Pearson M. L.,
 RA Lautenberger J. A., Papas T. S., Ghrayeb J., Chang N. T., Gallo R. C.,
 RA Wong-Staal F.;
 RT "Complete nucleotide sequence of the AIDS virus, HTLV-III";
 RL Nature 313:277-284(1985).
 XX
 RN [2]
 RA Muesing M. A., Smith D. H., Cabradilla C. D. JR., Benton C. V.,
 RA Kashy L. A., Capon D. J.;
 RT "Nucleic acid structure and expression of the human AIDS/
 RT lymphadenopathy retrovirus";
 RL Nature 313:450-458(1985).
 XX
 FH Key From To Description
 FH
 FT INVREP 1 2 inverted repeat
 FT SITE 1 634 long terminal repeat
 FT PBM 427 430 TATA-box
 FT SITE 453 453 U3 region
 FT CAP 454 454 cap site
 FT SITE 454 551 R region
 FT SITE 552 634 U5 region
 FT INVREP 633 634 inverted repeat
 FT SITE 635 653 tRNA binding site (tRNA-Lys)
 FT CDS 787 1182 gag p17
 FT CDS 787 2321 gag precursor polypeptide
 FT CDS 1183 2321 gag p24 and gag p15 for
 FT major capsid protein and for
 FT put. retroviral nucleic acid
 FT binding protein (NBP)(ref.2)
 FT (boundaries not defined)
 FT RPT 1968 2007 direct repeat
 FT RPT 2031 2065 direct repeat
 FT CDS 2007 5125 pol precursor polypeptides
 FT put. protease at 5' terminus
 FT reverse transcriptase
 FT put. endonuclease at 3' terminus
 FT RPT 2123 2163 direct repeat
 FT RPT 2164 2176 direct repeat
 FT CDS 3040 3648 SOR short open reading frame
 FT put. vestigial env gene
 FT CDS 6323 8821 env-lor precursor polypeptide
 FT CDS 6323 8821 envelope glycoprotein
 FT SITE 7736 7787 put. peptide cleavage site
 FT CDS 7757 8821 put. lor transmembrane
 FT protein

Sequence 9748 BP; 3431 A; 1781 C; 2358 G; 2168 T; 0 other;

X 10 20 30 40 50 60
ATGAGAGTGA-AGGAGAA-ATATCAGCACTTTGTGGAGA-TGGGGGTGGAAATGGGGGCAC-CATGCTCCTTGG
| | | | | | | | | | | | | | | | | | | | | |
CTAATAGAAAAGAGCAGAGGACAGTGGAAT----GAGAGTGAAGGAGAAATATCAGCACTTTGTGGAGATGGG
6230 6240 6250 6260 6270 6280 6290

[illegible]

130 140 150 160 170 180
 --CT--GTGTGGAA-----GGAGGCAA-CCACCA-CTCTATTTTGTGCATCAGATGCTAAAGCATATGAT
 || ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 TCTATTATGGCGTACCTGTGTGGGAAGGAAGCAACCACCACTCTATTTTG-----TGC---ATCAGATGCT
 6370 6380 6390 6400 6410 6420

180 200 210 220 230 240
 A---CAGAGG-TACATA-----AT--GTTTGGGCCACACATGCCTG--TGTAACCCACAGA-CCCCAACCCAC
 | | | | | | | | | | | | | | | |
 AAAGCATATGATACAGAGGTACATAATGTTTGGGC-CACA--CATGCCTGTGTACCCACAGACCCCAACC-C
 6430 6440 6450 6460 6470 6480 6490

250 260 270 280 290 300 310
 AAGCAGTATGATTGGTAAATGTGACAGAAAATTTTAACATGTGGAAAA--ATGACATGGTAGAACAGATG-C
 | | | | | | | | | | | | | | | | | |
 A-CGAGAAATA---GTA-TTG-----GTAAAT-GTGACA---GAAAATTTTAACAT-GTGGAA-AAATGAC
 5500 5510 5520 5530 5540 5550

[illegible]

350 400 410 420 430 440
 TAGT TTAA AGT GCACTGAT TTGG---GGAATGCTACTAAT---ACCAATACTAGTAATACCAATAGTAGTA
 : ||| | ||| || | ||| || || | ||| || |||||
 AAAATTAACCCCACTCTGTGTTAGT TTAAAGTGTC- ACTGATT TGAGAATGATACTAATACCAATAGTAGTA
 6620 6630 6640 6650 6660 6670 6680

450 460 470 480 490 500 510
 GCGGAGAAATGATGATGGAGAAAGGAGAGATAAAAAACTGCTCTTTCAATATCAGCACAAAGNATAAGAGGTA
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 GCGGAGAAATGATGATGGAGAAAGGAGAGATAAAAAACTGCTCTTTCAATATCAGCACAAAGCATAAGAGGTA
 650 670 6710 6720 6730 6740 6750

520 530 540 550 560 570 580
AGGTGCAGAAAGGATATGCATTTTCTTTATTAAGCTTCATATCATACCAATAGATAATGATACTACCAGCTATA
|||||
AGGTGCAGAGCGAGCTGTCATTTTCTTTTAAAGCTTCATATCATACCAATAGATAATGATACTACCAGCTATA

6720 6770 6780 6790 6800 6810 6820

590 600 610 620 630 640 650 660
CGTTGACAAAGTTGTAAACACCTCAATCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATAC
CGTTGACAAAGTTGTAAACACCTCAATCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATAC
6230 6240 6250 6260 6270 6280 6290

670 680 690 700 710 720 730
ATTATTTGTGCCCCGGCTGGTTTTGCGATTCTAAAATGTAATAATAAGACGTTCAATGGAACAGGACCATGTA
ATTATTTGTGCCCCGGCTGGTTTTGCGATTCTAAAATGTAATAATAAGACGTTCAATGGAACAGGACCATGTA
6900 6910 6920 6930 6940 6950 6960 6970

740 750 760 770 780 790 800
CAAATGTCAGCAGACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCA
CAAATGTCAGCAGACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCA
6980 6990 7000 7010 7020 7030 7040

810 820 830 840 850 860 870
GTCTAGCAGAAAGAGAGAGTAAATTAATCTGCGAATTTTCACAGACAATGCTAAAACCATTAATAGTACAGC
GTCTAGCAGAAAGAGAGAGTAAATTAATCTGCGAATTTTCACAGACAATGCTAAAACCATTAATAGTACAGC
7050 7060 7070 7080 7090 7100 7110

880 890 900 910 920 930 940
TGAACCAATCTGTAGAAATTAATTTGTACAGACCCCAACAACAATACAAGAAAAAGTATCCGTATCCAGAGGG
TGAACCAATCTGTAGAAATTAATTTGTACAGACCCCAACAACAATACAAGAAAAAGTATCCGTATCCAGAGAG
7120 7130 7140 7150 7160 7170 7180

950 960 970 980 990 1000 1010 1020
GACCAGGGAGAGCATTGTTTACAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAG
GACCAGGGAGAGCATTGTTTACAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAG
7190 7200 7210 7220 7230 7240 7250

1030 1040 1050 1060 1070 1080 1090
CAAAATGCAATGCCAGTTTAAACACAGATAGCTAGCAAATTAAGAGAACAAATTTGGAAATAATAAAACAATAA
CAAAATGCAATGCCAGTTTAAACACAGATAGCTAGCAAATTAAGAGAACAAATTTGGAAATAATAAAACAATAA
7260 7270 7280 7290 7300 7310 7320 7330

1100 1110 1120 1130 1140 1150 1160
TCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTTCT
TCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTTCT
7340 7350 7360 7370 7380 7390 7400

1170 1180 1190 1200 1210 1220 1230
ACTGTAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACA
ACTGTAATTCAACACAACTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTAAAGGGTCAAATAACA
7410 7420 7430 7440 7450 7460 7470

1240 1250 1260 1270 1280 1290 1300
CTGAAGGAAGTGACACAAATCACACTCCCATGCAGAATAAAACAAATTTATAACATGTGGCAGGAAGTAGGAA
CTGAAGGAAGTGACACAAATCACACTCCCATGCAGAATAAAACAAATTTATAACATGTGGCAGGAAGTAGGAA
7480 7490 7500 7510 7520 7530 7540

1310 1320 1330 1340 1350 1360 1370 1380
AAGCAATGTATCCCCCTCCCATCACTGACAGCAAAATTAGATGTTTCATCAAAATATTACAGGGCTGCTATTACAA
AAGCAATGTATCCCCCTCCCATCACTGACAGCAAAATTAGATGTTTCATCAAAATATTACAGGGCTGCTATTACAA

7550 7560 7570 7580 7590 7600 7610
1380 1400 1410 1420 1430 1440 1450
GAGATGGTGGTAATAACAACCATGGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGA
|||||
GAGATGGTGGTAATAACAACCATGGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGA
7620 7630 7640 7650 7660 7670 7680 7690
1450 1470 1480 1490 1500 1510 1520
GAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCCAAGAGAA
|||||
GAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCCAAGAGAA
7700 7710 7720 7730 7740 7750 7760
1530 1540 1550 1560 1570 1580 1590
GAGTGGTGGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCTTGGGTTCTTGGGAGCAGCAGGAA
|||||
GAGTGGTGGCAGAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCTTGGGTTCTTGGGAGCAGCAGGAA
7770 7780 7790 7800 7810 7820 7830
1600 1610 1620 1630 1640 1650 1660
GCACATATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGC
|||||
GCACATATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGC
7840 7850 7860 7870 7880 7890 7900
1670 1680 1690 1700 1710 1720 1730 1740
AGAACCAATTTCCTGAGGGCTATTEAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGC
|||||
AGAACCAATTTCCTGAGGGCTATTEAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGC
7910 7920 7930 7940 7950 7960 7970
1750 1760 1770 1780 1790 1800 1810
TCCAGGCAAGAACTCTGGCTGTGGAAAGATACCTAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTG
|||||
TCCAGGCAAGAACTCTGGCTGTGGAAAGATACCTAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTG
7980 7990 8000 8010 8020 8030 8040 8050
1820 1830 1840 1850 1860 1870 1880
GAAAACTCATTTCACCACTGCTGTGCCCTTGGAAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGA
|||||
GAAAACTCATTTCACCACTGCTGTGCCCTTGGAAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGA
8060 8070 8080 8090 8100 8110 8120
1890 1900 1910 1920 1930 1940 1950
ATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACATTCCTTAATTGAAG
|||||
ATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACATTCCTTAATTGAAG
8130 8140 8150 8160 8170 8180 8190
1960 1970 1980 1990 2000 2010 2020
AATCGCAAAACCAAGCAAGAAAGAAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATT
|||||
AATCGCAAAACCAAGCAAGAAAGAAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATT
8200 8210 8220 8230 8240 8250 8260
2030 2040 2050 2060 2070 2080 2090 2100
GGTTTAAACATACCAAAATTGGCTGTGGTATATAAAAATATTTCATAATGATAGTAGGAGGCTTGGTAGGTTTAA
|||||
GGTTTAAACATACCAAAATTGGCTGTGGTATATAAAAATATTTCATAATGATAGTAGGAGGCTTGGTAGGTTTAA
8270 8280 8290 8300 8310 8320 8330
2110 2120 2130 2140 2150 2160 2170
GAATAGTTTCTCTGTAATTTCTATAGTGAATAGAGTTAGGCAGGAGATATTCACCATTATCGTTTCAGACCC
|||||
GAATAGTTTCTCTGTAATTTCTATAGTGAATAGAGTTAGGCAGGAGATATTCACCATTATCGTTTCAGACCC

8340	8350	8360	8370	8380	8390	8400	8410
2180	2190	2200	2210	2220	2230	2240	
ACCTCCCAACCCCGAAGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACA							
ACCTCCCAATCCCGAAGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACA							
8420	8430	8440	8450	8460	8470	8480	
2250	2260	2270	2280	2290	2300	2310	
GATCGATTGCAATTAGTGAACCGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGCCTTGTGCCTCTTCAGC							
GATCGATTGCAATTAGTGAACCGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGCC-TGTGCCTCTTCAGC							
8490	8500	8510	8520	8530	8540	8550	
2320	2330	2340	2350	2360	2370	2380	
TACCAACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACCTCTGGGACGCAGGGGGTGGGAA							
TACCAACCGCTTGAGAGACTTACTCTTGATTGTAACGAGGATTGTGGAACCTCTGGGACGCAGGGGGTGGGAA							
8560	8570	8580	8590	8600	8610	8620	
2390	2400	2410	2420	2430	2440		
GCCCTCAAAATATTGCTGGGAATCTCCTACAGTATTGGAGTCAGGAACTAAAG							
GCCCTCAAAATATTGCTGGGAATCTCCTACAGTATTGGAGTCAGGAGCTAAAG							
8630	8640	8650	8660	8670	X		

XX
RN [3] (bases 508-5566) exons only, tat mRNA
RA Arya S.K., Guo C., Josephs S.F., Wong-staal F.;
RT "Trans-activator gene of human T-lymphotropic virus type III
RT (HTLV-III)";
RL Science 229:68-73(1985).

XX
RN [4] (bases 5775-6082, 8397-8499)
RA Sodroski J.G., Paterca R., Rosen C.A., Wong-staal F., Haseltine W.;
RT "Location of the trans-activating region on the genome of human
RT T-cell lymphotropic virus type III";
RL Science 229:74-77(1985).

XX
RN [5] mRNA splice sites
RA Rabson A.B., Daugherty D.F., Venkatesan S., Boulukos K.e.,
RA Benn S.I., Folks T.M., Feorino P., Martin M.;
RT "Transcription of novel open reading frames of AIDS retrovirus
RT during infection of lymphocytes";
RL Science 229:1388-1390(1985).

XX
RN [6] 27k antigen cds
RA Allan J.S., Coligan J.E., Lee T.H., McLane M.F., Kanki P.J.,
RA Gropman J.E., Essex M.;
RT "A new HTLV-III/LAV encoded antigen detected by antibodies from
RT AIDS patients";
RL Science 230:810-813(1985).

XX
RN [7] (bases 5778-8933) in hxb-3
RA Crowl R., Ganguly K., Gordon M., Conroy R., Schaber M., Kramer R.,
RA Shaw G., Wong-staal F., Reddy E.P.;
RT "HTLV-III env gene products synthesized in E.coli are recognized
RT by antibodies present in the sera of AIDS patients";
RL Cell 41:879-886(1985).

XX
RN [8] gp160 and gp120 coding sequences
RA Allan J.S., Coligan J.E., Barin F., McLane M.F., Sodroski J.G.,
RA Rosen C.A., Haseltine W.A., Lee T.H., Essex M.;
RT "Major glycoprotein antigens that induce antibodies in AIDS
RT patients are encoded by HTLV-III";
RL Science 228:1091-1094(1985).

XX
RN [9] regulatory sequences in the ltr
RA Rosen C.A., Sodroski J.G., Haseltine W.A.;
RT "The location of cis-acting regulatory sequences in the human T
RT cell lymphotropic virus type III (HTLV-III/LAV) long terminal
RT repeat";
RL Cell 41:813-823(1985).

XX
RN [10] (bases 1-8749)
RA Van Beveren C., Van Beveren C., Van Beveren C.;
RT "Appendix B: HTLV-3/LAV genome";
RL (in) Weiss R., Teich N., Varmus and Coffin J.M. (eds.);
RL RNA Tumor Viruses Second Edition:1102-1148
RL Cold Spring Harbor Laboratory, New York (1985)

XX
RN [11] trans-activator function and tar sequence
RA Rosen C.A., Sodroski J.G., Goh W.C., Dayton A.I., Lippke J.,
RA Haseltine W.A.;
RT "Post-transcriptional regulation accounts for the trans-activation
RT of the human T-lymphotropic virus type III";
RL Nature 318:555-559(1986).

XX
RN [12] pol coding sequence
RA Morzo Veronese F., Copeland T.D., DeVico A.L., Rahman R.,
RA Oroszlan S., Gallo R.C., Sarngedharan M.G.;
RT "Characterization of highly immunogenic p66/p51 as the reverse

RT transcriptase of HTLV-III/LAV";
 RL Science 231:1288-1291(1986).
 XX

RN [13] the 23k sor gene product
 RA Kan N.G. , Franchini G. , Wong-staal F. , DuBois G.C. , Robey W.G. ,
 RA Lautenberger J.A. , Papas T.S. ;
 RT "Identification of HTLV-III/LAV sor gene product and detection of
 RT antibodies in human sera";
 RL Science 231:1553-1555(1986).
 XX

RN [14] pol nh2-terminal region
 RA Kramer R.A. , Schaber M.D. , Skalka A.M. , Ganguly K. , Wong-staal F. ,
 RA Reddy E.P. ;
 RT "HTLV-III gag protein is processed in yeast cells by the virus
 RT pol-protease";
 RL Science 231:1580-1584(1986).
 XX

RN [15] sor 23K protein
 RA Lee T.H. , Colligan J.E. , Allan J.S. , McLane M.F. , Groopman J.E. ,
 RA Essex M. ;
 RT "A new HTLV-III/LAV protein encoded by a gene found in cytopathic
 RT retroviruses";
 RL Science 231:1546-1548(1986).
 XX

RN [16] sor 23k protein
 RA Sodroski J.G. , Goh W.C. , Rosen C.A. , Tartar A. , Portetelle D. ,
 RA Bunny A. , Haseltine W. ;
 RT "Replicative and cytopathic potential of HTLV-III/LAV with sor
 RT gene deletions";
 RL Science 231:1549-1553(1986).
 XX

RN [17] spi binding sites in the promoter region
 RA Jones K.A. , Kadonaga J.T. , Luciw P.A. , Tjian R. ;
 RT "Activation of the AIDS retrovirus promoter by the cellular
 RT transcription factor, Sp1";
 RL Science 232:755-759(1986).
 XX

RN [18] acceptor and donor splice sites for tat and 27k
 RA Arya S.K. , Gallo R.C. ;
 RT "Three novel genes of human T-lymphotropic virus type III: Immune
 RT reactivity of their products with sera from acquired immune
 RT deficiency syndrome patients";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:2209-2213(1986).
 XX

RN [19] deletion mutants in the tat gene
 RA Dayton A.L. , Sodroski J.G. , Rosen C.A. , Goh W.C. , Haseltine W.A. ;
 RT "The trans-activator gene of the human T cell lymphotropic virus
 RT type III is required for replication";
 RL Cell 44:841-847(1986).
 XX

RN [20] hypervariable and conserved regions in the env gene
 RA Willey R.W. , Ruthledge R.A. , Dias S. , Folks T. , Theodore T.S. ,
 RA Buckler C.E. , Martin M.A. ;
 RT "Identification of conserved and divergent domains within the
 RT envelope gene of the acquired immunodeficiency syndrom
 RT retrovirus";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5038-5042(1986).
 XX

RN [21] art cds boundaries
 RA Sodroski J.G. , Goh W.C. , Rosen C.A. , Dayton A. , Terwilliger E. ,
 RA Haseltine W. ;
 RT "A second post-transcriptional trans-activator gene required for
 RT HTLV-III replication";
 RL Nature 321:412-417(1986).
 XX

CC Sequence for [7] was kindly supplied in computer readable form by

R. Crowl, 09/17/85. R. Patarca provided site information and a clean copy for [4], 09/16/85. Acquired immune deficiency syndrome (AIDS) is caused by a retrovirus known by several names, perhaps representing two separate strains: human T-cell lymphotropic virus-III (HTLV-III), whose sequence is given below, and lymphadenopathy-associated virus (LAV) are thought to be one strain differing from AIDS-associated retrovirus type 2 (ARV-2) when overall homology is the criterion. Some reading frame similarities suggest that ARV-2 and LAV are more closely related. All three viruses, whose sequences do not differ by more than 6%, are believed to belong to the C type subfamily Lentiviridae, the "slow" retroviruses. The BH10 sequence differs from BH8 and BH5 by 0.9% in the coding regions and 1.9% in the noncoding regions, and the authors of [2] believe that these are stable variants. The 5' and 3' LTRs of BH10 and BH8 were not fully sequenced; the missing bases (493-675 and 5608-9749) were filled in by [2] from the proviral clone HXB2 [1]. The sequence below is that of BH10 with exception of the variation at position 9197 which allows annotation of the 27K coding sequence. The BH8 sequence spans bases 6033 to 9607, the BH5 sequence spans bases 675 to 6038, and the HXB3 sequence [7] spans bases 5778 to 8933. While this entry is offered as the reference locus for the AIDS retroviral sequence loci, no claim is being made that this sequence is more prevalent or typical than others, all of which have been entered in this library with annotation. The HTLV-III genome encodes at least six proteins or polyproteins: gag, pol, env, TAT, 27K antigen and the sor 23K product. The 3' ORF (positions 8797-9447) is truncated in BH10 (stop codon at positions 9196-9198), but reads through in BH8 and other sequences to yield what is now called the 27K antigen. The sequence below is from BH10 with exception of the variation at position 9197 which allows annotation of the 27K coding sequence. Additionally there are four short open reading frames, bases 1248-1406, 4442-4542, 5592-5828 and 6095-6340, which are conserved to a large degree. A seventh gene has been proposed based upon a combination of mutational and regulatory evidence: called "ART" (for anti-repression transactivator), its product appears to act post-transcriptionally to relieve negative repression of gag and env production [21]. The exon assignments for ART are putative, but if they are corroborated, the ART protein would be 116 amino acids in length. The mechanism for pol gene translation has not been elucidated; a gag-pol fusion protein is possible; splicing or frameshift have not been ruled out. The viral protease would be determined by the region in question. Approximately two-thirds of the variant sites in the gag and pol genes are "silent mutations", while over half of those in the env gene are not. Reference [20] defines divergent and conserved regions for the env gene. Because of the excessive variability of the env gene, differences between the sequences summarized herein and other env gene entries have not been annotated; only HTLV-III sequence variations have been included in the sites of this entry. Other entries will include information for alignment with this entry, including the Zaire and New York isolate sequences reported by [20]. The TAT protein (trans-activator protein, approximately 14 kd) is an effector of an autostimulatory pathway through interaction with a positive control element, the trans-activating responsive sequence, TAR. TAT seems to be a transcriptional control molecule in HTLV-I, but [11] demonstrates that it is a post-transcriptional regulatory molecule in HTLV-III. Deletion mutants in the TAT gene are incapable of prolific replication and exhibit no cytopathic effects in T4+ cell lines [19]. The TAR sequence(s) are found to be between -17 and +80 relative to the cap site +1 (base 455) and is highly conserved. Enhancer sequences which need not be viral-specific are found upstream from TAR [9],[11]. Three tandem decanucleotide Sp1 binding sites are located between bases 377 and 409, of which site III shows the strongest affinity for the cellular factor; intact, the three sites cause up to a tenfold effect on transcriptional

CC efficiency in vitro [117] (The authors demonstrate the existence of
 CC Sp1 in a human T-cell line). In addition to the ~9.4 kb genomic
 CC mRNA, subgenomic mRNAs of 7.4, 5.5, 5.0, 4.3, 2.0 and 1.8 have been
 CC detected. All are probably polyadenylated at the same site,
 CC position 9666 below, with a potential polyadenylation signal at
 CC 9642-9648, and capped at the same site, position 455, with a
 CC potential TATA box at 427-431. The doubly-spliced transcript of
 CC about 2.0 kb is responsible for the TAT message at least, and
 CC depending upon the acceptor site, also for the sor and 27K
 CC messages, given that a single, albeit partial, mRNA exists for all
 CC three [18]. The acceptor splice for TAT is at position 5811 and the
 CC putative acceptor splice for 27K is at position 6010; the donor
 CC splice site in all three cases would be at position 6079 [18]. The
 CC doubly spliced message would also encode the newly proposed ART
 CC protein.

XX				
FH	Key	From	To	Description
FH				
FT	RPT	1	634	5' LTR
FT	RPT	1	634	5' LTR
FT	VARIANT	82	82	a in BH10; g in H9
FT	VARIANT	101	101	g in BH10; a in H9
FT	VARIANT	108	108	a in [2], H9; g in HXB2 [1]
FT	VARIANT	154	154	g in [2]; t in HXB2 [1], H9
FT	VARIANT	168	168	t in [2]; g in HXB2 [1], H9
FT	VARIANT	176	176	a in [2]; g in HXB2 [1], H9
FT	VARIANT	183	183	c in [2], H9; t in HXB2 [1]
FT	VARIANT	227	227	a in [2], H9; g in HXB2 [1]
FT	VARIANT	291	291	a in [2]; g in HXB2 [1], H9
FT	VARIANT	333	333	c in [2]; t in HXB2 [1], H9
FT	SITE	377	386	Sp1 binding site III [17]
FT	SITE	388	397	Sp1 binding site II [17]
FT	SITE	399	408	Sp1 binding site I [17]
FT	VARIANT	421	421	c in BH10, BH5; t in H9
FT	RPT	454	551	R repeat 5' copy
FT	RPT	454	551	R repeat 5' copy
FT	PROVRL	454	9666	HTLV3 virion RNA
FT	CAP	455	455	genomic mRNA start (cap site) [10]
FT	CAP	455	455	TAT, ART mRNA exon 1 start (cap site) [10], [18], [21]
FT	VARIANT	501	501	a in BH10, BH5, H9; g in HXB2 [1]
FT	SITE	636	653	primer (Lys-tRNA) binding site
FT	VARIANT	654	654	c in BH10, BH5; t in H9
FT	VARIANT	677	677	g in BH10, BH5; ggag in H9
FT	VARIANT	704	704	tga in BH10, H9; g in BH5 [2]
FT	CDS	787	2325	gag polyprotein precursor
FT	VARIANT	1280	1280	a in BH10; g in BH5 [2], H9
FT	VARIANT	1431	1431	a in BH10; g in BH5 [2], H9
FT	VARIANT	1455	1455	t in BH10, H9; c in BH5 [2]
FT	VARIANT	1611	1611	a in BH10, H9; g in BH5 [2]
FT	VARIANT	1620	1620	c in BH10, H9; t in BH5 [2]
FT	VARIANT	1656	1656	a in BH10, H9; g in BH5 [2]
FT	VARIANT	1662	1662	t in BH10; c in BH5 [2], H9
FT	VARIANT	1673	1675	g in BH10, BH5; c in H9
FT	VARIANT	1722	1722	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1806	1806	g in BH10, BH5; a in H9
FT	VARIANT	1845	1845	a in BH10, BH5; g in H9
FT	VARIANT	1903	1903	a in BH10, H9; t in BH5 [2]
FT	VARIANT	1906	1906	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1923	1923	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1950	1950	g in BH10, H9; a in BH5 [2]
FT	VARIANT	1953	1953	g in BH10, H9; t in BH5 [2]
FT	VARIANT	1988	1988	c in BH10, H9; t in BH5 [2]
FT	VARIANT	1992	1992	c in BH10, H9; a in BH5 [2]
FT	VARIANT	2003	2003	g in BH10, H9; a in BH5 [2]
FT	VARIANT	2013	2013	g in BH10, H9; a in BH5 [2]

FT	CDS	2391	5125	pol polyprotein (NH2-terminus
FT				uncertain; AA at 2391)
FT	VARIANT	2468	2468	g in BH10; BH5; a in H9
FT	VARIANT	2591	2591	c in BH10; H9; t in BH5 [2]
FT	VARIANT	2600	2600	g in BH10; H9; a in BH5 [2]
FT	VARIANT	2741	2741	g in BH10; a in BH5 [2], H9
FT	VARIANT	2827	2827	a in BH10; H9; g in BH5 [2]
FT	VARIANT	2858	2858	a in BH10; H9; g in BH5 [2]
FT	VARIANT	2980	2980	c in BH10; H9; t in BH5 [2]
FT	VARIANT	3007	3007	tta in BH10; H9; gtg in BH5 [2]
FT	VARIANT	3097	3097	a in BH10; g in BH5 [2], H9
FT	VARIANT	3122	3122	c in BH10; H9; t in BH5 [2]
FT	VARIANT	3222	3222	c in BH10; H9; t in BH5 [2]
FT	VARIANT	3302	3302	ag in BH10; H9; ga in BH5 [2]
FT	VARIANT	3366	3366	g in BH10; H9; a in BH5 [2]
FT	VARIANT	3398	3398	g in BH10; BH5; a in H9
FT	VARIANT	3395	3395	c in BH10; H9; t in BH5 [2]
FT	VARIANT	3755	3755	a in BH10; BH5; g in H9
FT	VARIANT	3767	3767	g in BH10; H9; a in BH5 [2]
FT	VARIANT	3833	3833	t in BH10; BH5; c in H9
FT	VARIANT	3855	3855	t in BH10; BH5; c in H9
FT	VARIANT	3898	3898	c in BH10; BH5; t in H9
FT	VARIANT	3922	3922	a in BH10; H9; g in BH5 [2]
FT	VARIANT	3934	3934	a in BH10; BH5; g in H9
FT	VARIANT	3954	3954	g in BH10; BH5; c in H
FT	VARIANT	3962	3962	caa in BH10; H9; tag in BH5 [2]
FT	VARIANT	3977	3977	g in BH10; H9; a in BH5 [2]
FT	VARIANT	3984	3984	c in BH10; H9; a in BH5 [2]
FT	VARIANT	3993	3993	a in BH10; H9; c in BH5 [2]
FT	VARIANT	4010	4010	a in BH10; g in BH5 [2], H9
FT	VARIANT	4016	4016	g in BH10; H9; a in BH5 [2]
FT	VARIANT	4029	4029	t in BH10; H9; c in BH5 [2]
FT	VARIANT	4049	4049	a in BH10; g in BH5 [2], H9
FT	VARIANT	4064	4064	c in BH10; H9; t in BH5 [2]
FT	VARIANT	4116	4116	a in BH10; BH5; c in H9
FT	VARIANT	4167	4167	g in BH10; BH5; c in H9
FT	VARIANT	4292	4292	t in BH10; H9; a in BH5 [2]
FT	CDS	5074	5652	sor 23K protein
FT	VARIANT	5156	5156	a in BH10; H9; g in BH5 [2]
FT	VARIANT	5314	5314	t in BH10; BH5; c in H9
FT	VARIANT	5348	5348	a in BH10; H9; g in BH5 [2]
FT	VARIANT	5401	5401	t in BH10; H9; c in BH5 [2]
FT	VARIANT	5412	5412	c in BH10; H9; t in BH5 [2]
FT	VARIANT	5548	5548	a in BH10; H9; g in BH5 [2]
FT	VARIANT	5628	5628	g in BH10; H9; a in BH5 [2]
FT	VARIANT	5846	5846	g in BH10; H9; HXB3; a in BH5 [2]
FT	CDS	5864	6078	TAT protein, exon 2 (first expressed
FT				exon)
FT	VARIANT	5934	5934	a in BH10; H9; HXB3; c in BH5 [2]
FT	CDS	6003	6078	ART protein, exon 2 (first expressed
FT				exon; putative)
FT	VARIANT	6035	6045	cctcctcaagg in BH10, HXB3 [7];
FT				gctcatcgaag in BH8 [2];
FT				g in BH5 [2], clone 12 cDNA [21]
FT	VARIANT	6086	6086	g in BH10, BH8, H9; a in HXB3 [7]
FT	VARIANT	6086	6086	t in BH10, HXB3 [7], H9; c in BH8 [2]
FT	VARIANT	6108	6108	a in BH10, HXB3 [7], H9; c in BH8 [2]
FT	VARIANT	6113	6114	gc in BH10, HXB3 [7], H9;
FT				gtaac in BH8 [2]
FT	VARIANT	6124	6124	a in BH10, HXB3 [7], H9; c in BH8 [2]
FT	VARIANT	6152	6152	g in BH10, HXB3 [7], BH8; c in H9
FT	CDS	6255	8325	envelope protein precursor (env)
FT	VARIANT	6373	6373	a in BH10, HXB3 [7], H9; t in BH8 [2]
FT	VARIANT	6474	6474	t in BH10, BH8 [2], H9; g in HXB3 [7]
FT	VARIANT	6748	6748	t in BH10, HXB3 [7], H9; a in BH8 [2]
FT	VARIANT	6809	6829	t in BH10, HXB3 [7], H9; c in BH8 [2]

FT	VARIANT	7088	7088	a in BH10; H9; g in BH8 [2], HXB3 [7]
FT	VARIANT	7119	7119	a in BH10; HXB3 [7], H9; g in BH8 [2]
FT	VARIANT	7121	7123	cca in BH10, H9; cac in BH8 [2],
FT				HXB3 [7]
FT	VARIANT	7171	7172	gt in BH10, H9; aa in BH8 [2], HXB3 [7]
FT	VARIANT	7187	7187	a in BH10, H9; g in BH8 [2], HXB3 [7]
FT	VARIANT	7272	7273	aa in BH10, H9; gc in BH8 [2], HXB3 [7]
FT	VARIANT	7291	7291	a in BH10, BH8 [2], H9; c in HXB3 [7]
FT	VARIANT	7343	7343	g in BH10, BH8 [2]; a in HXB3 [7], H9
FT	VARIANT	7439	7454	gtttaatagtagtctgg in BH10, HXB3 [7],
FT				and H9
FT	VARIANT	7461	7461	a in BH10, BH8 [2]; g in HXB3 [7], H9
FT	VARIANT	7493	7493	c in BH10, BH8 [2]; a in HXB3 [7], H9
FT	VARIANT	7521	7521	a in BH10, BH8 [2]; t in HXB3 [7], H9
FT	VARIANT	7574	7574	t in BH10, CH8 [2]; c in HXB3 [7], H9
FT	VARIANT	7636	7636	g in BH10, BH8 [2]; a in HXB3 [7], H9
FT	VARIANT	7636	7637	cg in BH10, HXB3 [7], H9; gc in BH8 [2]
FT	VARIANT	7645	7645	a in BH10, BH8 [2], H9; g in HXB3 [7]
FT	VARIANT	8080	8061	ca in BH10, BH8 [2], H9; ac in H
FT	VARIANT	8127	8127	a in BH10, BH8 [2], H9; c in HXB [7]
FT	VARIANT	8131	8131	t in BH10, BH8 [2], H9; c in HXB3 [7]
FT	VARIANT	8135	8135	c in BH10, BH8 [2], H9; g in HXB3 [7]
FT	VARIANT	8257	8257	g in BH10, BH8, HXB3; a in H9
FT	VARIANT	8273	8273	t in BH10, BH8, HXB3; g in H9
FT	VARIANT	8364	8364	g in BH10, HXB3 [7]; a in BH8 [2], H9
FT	CDS	8409	8454	TAT protein, exon 3 (AA at 8410)
FT	CDS	8409	8663	ART protein, exon 3 (putative; AA at
FT				8411)
FT	VARIANT	8422	8422	t in BH10, HXB3 [7], clone 12 cDNA
FT				[21]; a in BH8 [2]; c in H9
FT	VARIANT	8464	8464	g in BH10, BH8, HXB3, clone 12 cDNA
FT				[21]; a in H9
FT	VARIANT	8657	8657	g in BH10, BH8 [2]; a in HXB3 [7], H9,
FT				clone 12 cDNA [21]
FT	VARIANT	8672	8672	g in BH10, HXB3 [7], clone 12 cDNA
FT				[21], H9; a in BH8 [2]
FT	VARIANT	8692	8692	g in BH10, HXB3 [7], clone 12 cDNA
FT				[21], H9; a in BH8 [2]
FT	VARIANT	8748	8748	g in BH10, HXB3 [7], clone 12 cDNA
FT				[21], H9; t in BH8 [2]
FT	VARIANT	8758	8758	g in BH10, H9; c in BH8 [2];
FT				a in HXB3 [7], clone 12 cDNA [21]
FT	VARIANT	8771	8771	t in BH10, HXB3 [7], clone 12 cDNA
FT				[21], H9; c in BH8 [2]
FT	CDS	8827	9447	27K protein, exon 3 (first expressed
FT				exon)
FT	VARIANT	8857	8857	g in BH10, BH8, HXB3, clone 12 cDNA
FT				[21]; a in H9
FT	VARIANT	8924	8924	c in BH10, HXB3 [7], clone 12 cDNA
FT				[21], H9; t in BH8 [2]
FT	VARIANT	8967	8967	c in BH10, clone 12 cDNA [21], H9;
FT				t in BH8 [2]
FT	VARIANT	8978	8978	a in BH10, clone 12 cDNA [21], H9;
FT				c in BH8 [2]
FT	VARIANT	8985	8985	t in BH10, clone 12 cDNA [21], H9;
FT				c in BH8 [2]
FT	VARIANT	8987	8987	a in BH10, BH8; c in H9, clone 12
FT				cDNA [21]
FT	VARIANT	8994	8994	c in BH10, clone 12 cDNA [21], H9;
FT				t in BH8 [2]
FT	VARIANT	9019	9019	g in BH10, BH8; a in H9, clone 12
FT				cDNA [21]
FT	RPT	9116	9749	3' LTR
FT	VARIANT	9163	9166	t in BH10, clone 12 cDNA [21];
FT				c in BH8 [2]
FT	VARIANT	9187	9197	a in BH8 [2], H9, clone 12 cDNA [21];

FT				a in BH10 [2]
FT	VARIANT	9218	9218	g in BH10, BH8; a in H9, clone 12
FT				cdna [21]
FT	VARIANT	9212	9223	ga in BH10; clone 12 cdna [21]; H9;
FT				ag in BH8[2]
FT	VARIANT	9273	9279	g in BH10; BH8; clone 12 cdna [21];
FT				t in H9
FT	VARIANT	9283	9283	t in BH10; BH8; clone 12 cdna [21];
FT				g in H9
FT	VARIANT	9284	9284	t in BH10; H9; clone 12 cdna [21];
FT				a in BH8 [2]
FT	VARIANT	9291	9291	a in BH10; BH8; clone 12 cdna [21];
FT				g in H9
FT	VARIANT	9297	9297	c in BH10; clone 12 cdna [21]; H9;
FT				t in BH8 [2]
FT	VARIANT	9354	9354	g in BH10; HIVDSM); H9; t in BH8 [2]
FT	VARIANT	9406	9406	a in BH10; BH8; g in H9, clone 12 cdna
FT				[21]
FT	VARIANT	9448	9448	c in BH10; t in BH8 [2]; H9, clone 12
FT				cdna
FT	VARIANT	9536	9563	c in BH10; BH8; clone 12 cdna [21];
FT				g in H9
FT	RPT	9570	9666	R repeat 3' copy
FT	VARIANT	9616	9616	g in HXB2; a in H9, clone 12 cdna [21]
FT	VARIANT	9621	9621	g in HXB2; a in H9, clone 12 cdna [21]
FT	VARIANT	9663	9663	t in BH10; H9; tg in clone 12 cdna
FT				[21]
FT	POLYA	9666	9666	TAT; ART; 27K mRNA exon 3 end
FT				(poly-A site) [10],[18],[21]
FT	POLYA	9666	9666	genomic mRNA end (poly-A site) [10]
XX				
SQ	Sequence	5749 BP:	3431 A; 1781 C; 2369 G; 2168 T; 0 other;	

Initial Score = 1872 Optimized Score = 2176 Significance = 0.00
Residue Identity = 89% Matches = 2243 Mismatches = 150
Gaps = 106 Conservative Substitutions = 0

X	10	20	30	40	50	60
ATGAGAGTGA-AGGAGAA-ATATCAGCACTTGTGGAGA-TGGGGGTGGAAATGGGGGCAC-CATGCTCCTTGG						
CTGATAGAAAGAGAGCAAAAGACAGTGGCAAT-----GAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGG						
6230	6240	6250	6260	6270	6280	6290
70	80	90	100	110	120	
GATATTGATG-----AT-CT-GTAGTGCTACAGAAAAATTGT-GGGTCACAG-TCTATTATGGGGTAC--						
GGTGGAGATGGGGCACCATGCTCCTTGGGATGTTGATGATCTGTAGTGCTACAGAAAAATTGTGGGGTCACAG						
6300	6310	6320	6330	6340	6350	6360
130	140	150	160	170	180	
-CT---GTGTGAA-----GGAAGCAA-CCACCA-CTCTATTTTGTGCATCAGATGCTAAAGCATATGAT						
TCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCACCCTCTATTTTG-----TGC---ATCAGATGCT						
6370	6380	6390	6400	6410	6420	
190	200	210	220	230	240	
A---CAGAGG-TACATA-----AT--GTTTGGGCCACACATGCCTG--TGTACCCACAGA-CCCCAACCCAC						
AAAGCATAAGATACAGAGGTACATAATGTTTGGGC-CACA--CATGCCTGTGTACCCACAGACCCCAACC-C						
6430	6440	6450	6460	6470	6480	6490
250	260	270	280	290	300	310
AAGAGGTATGTTTGGTAAATGTGACAGAAAATTTTAAACATGTGGAAAA--ATGACATGGTAGAACAGATG-C						
A-CAGAAATGTA--GTA-TTG-----GTAAAT-GTGACA-----GAAAATTTTAAACAT-GTGGAA-AAATGAC						
6500	6510	6520	6530	6540	6550	

320	330	340	350	360	370	380
ATGAGGATATAATCAG-TTTTATE-EGATCAAAGCCTAAAGCCATGTG-TAAAA--TTAACCCCACTCTGTGT						
AT---GGTAGAA-CAGATGCATGAGGATATAA---TCAGTTTATGGGATCAAAGCCTAAAGCCA---TGTGT						
6560	6570	6580	6590	6600	6610	

390	400	410	420	430	440	
TAGTTTAA-AGTGCACCTGATTTGG---GGAATGCTACTAAT---ACCAATACTAGTAATACCAATAGTAGTA						
AAAATTAACCCCACTCTGTGTTAGTTTAAAGTGC-ACTGATTTGAAGAATGATACTAATACCAATAGTAGTA						
6620	6630	6640	6650	6660	6670	6680

450	460	470	480	490	500	510
GCGCGGAATGATGATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAGNATAAGAGGGTA						
GCGCGGAATGATGATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAGCATAGAGGGTA						
6690	6700	6710	6720	6730	6740	6750

520	530	540	550	560	570	580
AGGTTCAGAAAGAAATATGCATTTTTTTTATAAACTTGATATAATACCAATAGATAATGATACTACCAGCTATA						
AGGTTCAGAAAGAAATATGCATTTTTTTTATAAACTTGATATAATACCAATAGATAATGATACTACCAGCTATA						
6760	6770	6780	6790	6800	6810	6820

590	600	610	620	630	640	650	660
CGTTGACAAAGTTTGTACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATAC							
CGTTGACAAAGTTTGTACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATAC							
6830	6840	6850	6860	6870	6880	6890	6900

670	680	690	700	710	720	730
ATTATTTGTGCCCCGGCTGGTTTTTCGATTCTAAAATGTAATAATAAGACGTTCAATGGAACAGGACCATGTA						
ATTATTTGTGCCCCGGCTGGTTTTTCGATTCTAAAATGTAATAATAAGACGTTCAATGGAACAGGACCATGTA						
6910	6920	6930	6940	6950	6960	6970

740	750	760	770	780	790	800
CAAAATGTCAGCAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCA						
CAAAATGTCAGCAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCA						
6980	6990	7000	7010	7020	7030	7040

810	820	830	840	850	860	870
GTCTTGCCTGAAGGAGAGGTAGTAATTAGATCTGCCAATTTACAGACAATGCTAAAACCATAATAGTACAGC						
GTCTTGCCTGAAGGAGAGGTAGTAATTAGATCTGCCAATTTACAGACAATGCTAAAACCATAATAGTACAGC						
7050	7060	7070	7080	7090	7100	7110

880	890	900	910	920	930	940
TGAAUCCAATCTGTAGAAATTAAATTGTACAGAGACCCCAACAACAATACAAGAAAAAGTATCCGTATCCAGAGGG						
TGAAUCCAATCTGTAGAAATTAAATTGTACAGAGACCCCAACAACAATACAAGAAAAAGTATCCGTATCCAGAGAG						
7120	7130	7140	7150	7160	7170	7180

950	960	970	980	990	1000	1010	1020
GACCAAGGAGAGACATTTGTTACAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAG							
GACCAAGGAGAGACATTTGTTACAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAG							
7190	7200	7210	7220	7230	7240	7250	7260

1030	1040	1050	1060	1070	1080	1090
CAAAATGCAATGACATTTTAAAGACAGATAGCTAGCAAATTAAGAGAGACAATTTGGAAATAATAAAACAATAA						
CAAAATGCAATGACATTTTAAAGACAGATAGCTAGCAAATTAAGAGAGACAATTTGGAAATAATAAAACAATAA						
7270	7280	7290	7300	7310	7320	7330

1100	1110	1120	1130	1140	1150	1160
TCTTTAAGCAATGCTCAGGCGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTTCT						
TCTTTAAGCAATGCTCAGGCGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTTCT						
7340	7350	7360	7370	7380	7390	7400

1170	1180	1190	1200	1210	1220	1230
ACTGTAAATGCAACAGAACTGTTTAATACTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACA						
ACTGTAAATGCAACAGAACTGTTTAATACTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAAATAACA						
7410	7420	7430	7440	7450	7460	7470

1240	1250	1260	1270	1280	1290	1300
CTGAAGGAAGTGCACAAATCAGACTCCCATGCAGAAATAAACAAATTTATAACATGTGGCAGGAAGTAGGAA						
CTGAAGGAAGTGCACAAATCAGACTCCCATGCAGAAATAAACAAATTTATAACATGTGGCAGGAAGTAGGAA						
7480	7490	7500	7510	7520	7530	7540

1310	1320	1330	1340	1350	1360	1370	1380
AAGCAATGTATGCCCCGCCATCAGCGGACAAATTAGATGTTTCATCAAATATTACAGGGCTGCTATTACAA							
AAGCAATGTATGCCCCGCCATCAGCGGACAAATTAGATGTTTCATCAAATATTACAGGGCTGCTATTACAA							
7550	7560	7570	7580	7590	7600	7610	7620

1390	1400	1410	1420	1430	1440	1450
GAGAGGTGGTCAATACAAACATGGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGA						
GAGAGGTGGTCAATACAAACATGGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGA						
7630	7640	7650	7660	7670	7680	7690

1460	1470	1480	1490	1500	1510	1520
GAAGTGAATTATATATATATAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAA						
GAAGTGAATTATATATATATAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAA						
7700	7710	7720	7730	7740	7750	7760

1530	1540	1550	1560	1570	1580	1590
GAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTTCCTTGGGTTCTTGGGAGCAGCAGGAA						
GAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTTCCTTGGGTTCTTGGGAGCAGCAGGAA						
7770	7780	7790	7800	7810	7820	7830

1600	1610	1620	1630	1640	1650	1660
GCATATGGGCGCAGCTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGC						
GCATATGGGCGCAGCTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGC						
7840	7850	7860	7870	7880	7890	7900

1670	1680	1690	1700	1710	1720	1730	1740
AGAAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGC							
AGAAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGC							
7910	7920	7930	7940	7950	7960	7970	7980

1750	1760	1770	1780	1790	1800	1810
TCCAGGCAAGAAATCCTGGCTGTGAAAGATACCTAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTG						
TCCAGGCAAGAAATCCTGGCTGTGAAAGATACCTAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTG						
7990	8000	8010	8020	8030	8040	8050

1820	1830	1840	1850	1860	1870	1880
GAAAGACTCATTTGCAAGCACTGCTGTGCCTTGGAAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGA						
GAAAGACTCATTTGCAAGCACTGCTGTGCCTTGGAAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGA						
8060	8070	8080	8090	8100	8110	8120

[illegible]

[illegible]

```

XX
AC      X04414;
XX
DT      17-OCT-1986 (incorporated)
XX
DE      Human lymphadenopathy virus (ELI isolate), complete genome.
XX

```


500 700 800 900 1000 1010 1020
GTTGCGGCACTTCTGACGATATGAAATGCGGATATGAGACACAGCATTGTAACATTAGTAGAGCAAAAT

[illegible]

7430 7500 7510 7520 7530 7540 7550 7560
 1820 1830 1840 1850 1860 1870 1880
 AAACACATTTTCACACGCTTCGAGCTTGGAGTAATAAATCTCTGGAACAGATTGGAAT
 7570 7580 7590 7600 7610 7620 7630
 1890 1900 1910 1920 1930 1940 1950
 AACACACATTTTCACACGCTTCGAGCTTGGAGTAATAAATCTCTGGAACAGATTGGAAT
 7640 7650 7660 7670 7680 7690 7700
 1960 1970 1980 1990 2000 2010 2020 2030
 TACACACATTTTCACACGCTTCGAGCTTGGAGTAATAAATCTCTGGAACAGATTGGAAT
 7710 7720 7730 7740 7750 7760 7770
 2040 2050 2060 2070 2080 2090 2100
 TACACACATTTTCACACGCTTCGAGCTTGGAGTAATAAATCTCTGGAACAGATTGGAAT
 7780 7790 7800 7810 7820 7830 7840
 2110 2120 2130 2140 2150 2160 2170
 ATACATTTTCACACGCTTCGAGCTTGGAGTAATAAATCTCTGGAACAGATTGGAAT
 7850 7860 7870 7880 7890 7900 7910 7920
 2180 2190 2200 2210 2220 2230 2240
 CTACACATTTTCACACGCTTCGAGCTTGGAGTAATAAATCTCTGGAACAGATTGGAAT
 7930 7940 7950 7960 7970 7980 7990
 2350 2360 2370 2380 2390 2400 2410
 TACACATTTTCACACGCTTCGAGCTTGGAGTAATAAATCTCTGGAACAGATTGGAAT
 8000 8010 8020 8030 8040 8050 8060
 2320 2330 2340 2350 2360 2370 2380 2390
 CCACACATTTTCACACGCTTCGAGCTTGGAGTAATAAATCTCTGGAACAGATTGGAAT
 8070 8080 8090 8100 8110 8120 8130
 2400 2410 2420 2430 2440
 CCTACATTTTCACACGCTTCGAGCTTGGAGTAATAAATCTCTGGAACAGATTGGAAT
 8140 8150 8160 8170 8180 X

5. KUNZ-158-CL33 SER

RENTLV4C HTLV-3 (HTLV-4) partial provirus DNA (lambda PKE10)

ID RENTLV4C standard DNA: 5391 BP.
 XX
 AC K05391.1 V002531
 XX
 DT 24 MAR 1985 (annotation)

XX
DE HTLV-3 (HTLV-4) partial provirus DNA (lambda PKE102)
XX
KW env gene; gag gene; long terminal repeat; overlapping genes;
KW unidentified reading frame.
XX
OS Human T-cell leukemia virus type IV
OC Viridae; ssRNA enveloped viruses; Retroviridae.
XX
RN [1] (bases 1-5031)
RA Hahn B.H. ;
RT ;
RL Submitted (09-DEC-1987) on tape to the EMBL Data Library by:
RL Hahn B.H. , University of Alabama at Birmingham (UAB), 1918
RL University Blvd, BHSB 208, Birmingham, AL 35295.
XX
RN [2]
RA Hahn B.H. ; Kong L.L. ; Lee S.W. ; Kumar R. ; Taylor M.E. ; Arya S.K. ,
RA Shaw G.H. ;
RT "Isolation of HTLV-4 to Simian and human
RT immunodeficiency-associated viruses";
RL Nature 330:194-196(1987).
XX
CC -source: clone= lambda PKE 102
CC lambda PKE 102 is derived from PK82(HTLV-IV);
CC it is assumed that this isolate is a laboratory contaminants
CC of SIV mac 251.

XX
FH Key From To Description
FH
FT CDS 1 227 ser (1 is 2nd base in codon)
FT CDS 59 393 X (AA 1-112)
FT CDS 357 687 R (AA 1-97)
FT CDS 850 3452 env (AA 1-881)
FT CDS 3328 4117 3'ORF (AA 1-263)
FT RTT 3714 4521 long terminal repeat
FT CDS 4755 51391 gag (AA 1-212)
XX
SQ Sequence 5551 bp; 1732 A; 1035 C; 1378 G; 1242 T; 0 other;

Initial Score = 298 Optimized Score = 1247 Significance = 0.00
Residue Identity = 55% Matches = 1429 Mismatches = 889
Gaps = 266 Conservative Substitutions = 0

```

X      10      20      30      40      50      60
ATGAGAGTGAGGAGAGAAATAACAGCACTT--GTGGAGATGGG----GGTGGA--AATGGGGGCACCATGCTC
||  ||      |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
ATCACTGCTTATCGCCATCTT-TGCTTTTAAGTGTCTATGGGATCTATTGTACTCAATATGTCA-CAGTCT-
X 970      890      990      900      910      920      930

      70      80      90      100      110      120      130
CTTGGGATATTGATGATCTGTAG--TGCTACAGAAAAATTGTGGGTCACAGTCTATTATGGGGGTACCTGTG
||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
TTTATGGTGTATACAGCT-TGAGGAATGCGACAATTCCCCTCTTCTGTGCAACCAAGAATAGGGATACT-TG
940      950      960      970      980      990      1000

      140      150      160      170      180      190
TGGAGGAGAGCA--AGC-ACCACCTCTATTTTG-TGCATCAGATGCTAAAGCAT-ATGATACAGAGGTACATA
||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
GGGAGCAATCAATGCTACAGATTAATGGTGATTATTTCAGA-ATTGGCCCTTAATGTTACAGAAAGCTTTG
1010      1020      1030      1040      1050      1060      1070

200      210      220      230      240      250      260      270
ATGTTTGGGGCAGACATGCGCTGTGTACCCACAGACCCCAACCCACAGAAGTAGTATTGGTA-AATGTGACA
||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
CTACTGGGAGAG--CT-ACAGT-----CACAGAGC--AGGCAATAGAGGACGTATGGCAACTCTTTGAGA

```

1080 1090 1100 1110 1120 1130

240 250 300 310 320 330

-GAAATTTTAAACATGTEGAAAA--ATGA-CAT--GGTAGAAC----AGATGCATGAGGATATAATCAG-TT

111 1111 1111 11 111 11 11 11111 11 11 1111

1140 1150 1160 1170 1180 1190 1200

340 350 360 370 380 390

TAT-EGGAT--CA-AGGCCTAAA-GCCATGTGTATAAATTAACCCC-AC-TCTGTGTTAGT-TTAAAGTGCA

11 1111 11 1111 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

1210 1220 1230 1240 1250 1260 1270

400 410 420 430 440 450 460

CTGATTTGGGGAATGCTACTAATAACCAATACTAG-TAATACCAATAGTAGTAGCGGGGAA-ATGA-TGATGG

11 111 1111 1111 11 11 111 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

1280 1290 1300 1310 1320 1330 1340 1350

470 480 490 500 510 520 530

AGGAGGAGAGGCTAAAAAAGTCTCTTTCAATATCAGCACAGNATAAGAGGTAAGGTGCAGAAAG--AATA

11 11

1360 1370 1380 1390 1400 1410 1420

540 550 560 570 580 590 600

TGC--ATTT-TTATATAAAGTTGATATAATACCAATAGATAATGATACTACCAGCTATACGTTGACAAGTTG

11 1111 111 11 1111 11 11111111 11 11 1111 11 11 11 11 11 11 11 11

1430 1440 1450 1460 1470 1480

610 620 630 640 650 660 670

TAAACCTTCACTCATTACACAGGCCTGT-CCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCC

11111 11 11 11 11 11 1111 11 11 11 11 11 11 11 11 11 11 11 11 11 11

1490 1500 1510 1520 1530 1540 1550

680 690 700 710 720 730 740

CGGCTGGT TTTGGCATCTTAHAATGTATAATA-AGACGTTCAATGGAACAGGAC--CATGTACAAATGT-C

11 1111 111 11 11111111 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

1560 1570 1580 1590 1600 1610 1620 1630

750 760 770 780 790 800

AGCAGAGTACAAATGTACACATGGA--AT-TAGGC-CAGTAGTATCAACTCAACTGCTGTTGAATGGCAGTCT

11 11 11 111 1111 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

1640 1650 1660 1670 1680 1690

810 820 830 840 850 860 870

AGCAGAGAGT-AGAGGTAGTATTAGATCTGCCAATTTACAGACAATGCTAAAACCATAA-TAG--TACAGC

111111 111 11 1111 1111 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

1700 1710 1720 1730 1740 1750 1760

880 890 900 910 920 930 940

TGAACCAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGTATCCGTATCCAGAGGG

11 11 11 11 11 11 11111111 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

1770 1780 1790 1800 1810 1820

950 960 970 980 990 1000 1010

GACCA--GGGAGAGCATT--TGTACAATAGGAAAAATAGGAAATATGAGAC-AAGCA--CATTGTAACATT

1111 11 111 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

CACCATTTATGCTGGATTGGTTTTCCACTCACACCA-ATCAATGATAGGCCAAAGCAGGCATGGT--GT

1830 1840 1850 1860 1870 1880

1020 1030 1040 1050 1060 1070

AG--TAGAGCAAGATGCAATGCCACTTTAAACAGAT-AGCTAGCAAAATT----AAGAGAACA---AT-TTG

1890 1900 1910 1920 1930 1940 1950 1960

1080 1090 1100 1110 1120 1130 1140

GAAATAATAAAC-AAT--AATC---TTTAAAGCAATCCT-CAGGAGGGGGACCCAGAAATTGTAACGCACA-GT

1970 1980 1990 2000 2010 2020

1150 1160 1170 1180 1190 1200

---TTTAATGTGTGAGGGGAATTTTTCTACTGTAATTCAACACAACCTGTTTAATAGTACTTGGTTTA--ATAG

2030 2040 2050 2060 2070 2080 2090

1210 1220 1230 1240 1250 1260 1270 1280

TACTTGGAGTACTGAAGGGTCAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCAGAATAAAACA

2100 2110 2120 2130 2140 2150 2160

1290 1300 1310 1320 1330 1340 1350

ATTATAAACATGTGGCAGGAAGTAGGAAAAGCAATG--TATGCCCTCCCATCAGCGGACAAATTAGATGT

2170 2180 2190 2200 2210 2220 2230

1360 1370 1380 1390 1400 1410 1420

TCATCAAAATTATACAGGGCTGCTATTAAACAAGAGATGGTGGTAATAACAACAATGGGTCCGAGATCTTCAGA

2240 2250 2260 2270 2280 2290 2300

1430 1440 1450 1460 1470 1480 1490

CCTG--GAGGAGG-AGATATGAGGGACAAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAA

2310 2320 2330 2340 2350 2360

1500 1510 1520 1530 1540 1550

CCATTAGGAGTATACACCCACCAAGGCAGAGAG----AAGAGTGGT-GCA--GAGAGAAAAAGAGCAGTGGG

2370 2380 2390 2400 2410 2420 2430

1560 1570 1580 1590 1600 1610 1620

AATAGGAGCTTTGTTCTTTGGTTCT-TGGGAGCAGCAGGAAGCACTATGGGCGCACGGTCAATGACGCTGA

2440 2450 2460 2470 2480 2490 2500

1630 1640 1650 1660 1670 1680 1690

CGGTACAGGCGAGACAATTATTGTGTATAGTGACAGCAGCAGAACAAATTTGCTGAGGGCTATTGAGGCGC

2510 2520 2530 2540 2550 2560 2570

1700 1710 1720 1730 1740 1750 1760 1770

AAGAGCATGTGTGCAACTCAGAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGAT

2580 2590 2600 2610 2620 2630 2640 2650

2660 2670 2680 2690 2700 2710 2720 2730

2740 2750 2760 2770 2780 2790 2800 2810

2820 2830 2840 2850 2860 2870 2880 2890

2900 2910 2920 2930 2940 2950 2960 2970

2980 2990 3000 3010 3020 3030 3040 3050

3060 3070 3080 3090 3100 3110 3120 3130

3140 3150 3160 3170 3180 3190 3200 3210

3220 3230 3240 3250 3260 3270 3280 3290

3300 3310 3320 3330 3340 3350 3360 3370

3380 3390 3400 3410 3420 3430 3440 3450

3460 3470 3480 3490 3500 3510 3520 3530

3540 3550 3560 3570 3580 3590 3600 3610

3620 3630 3640 3650 3660 3670 3680 3690

3700 3710 3720 3730 3740 3750 3760 3770

3780 3790 3800 3810 3820 3830 3840 3850

3860 3870 3880 3890 3900 3910 3920 3930

3940 3950 3960 3970 3980 3990 4000 4010

4020 4030 4040 4050 4060 4070 4080 4090

4100 4110 4120 4130 4140 4150 4160 4170

4180 4190 4200 4210 4220 4230 4240 4250

4260 4270 4280 4290 4300 4310 4320 4330

4340 4350 4360 4370 4380 4390 4400 4410

4420 4430 4440 4450 4460 4470 4480 4490

4500 4510 4520 4530 4540 4550 4560 4570

4580 4590 4600 4610 4620 4630 4640 4650

4660 4670 4680 4690 4700 4710 4720 4730

4740 4750 4760 4770 4780 4790 4800 4810

4820 4830 4840 4850 4860 4870 4880 4890

4900 4910 4920 4930 4940 4950 4960 4970

4980 4990 5000 5010 5020 5030 5040 5050

5060 5070 5080 5090 5100 5110 5120 5130

5140 5150 5160 5170 5180 5190 5200 5210

5220 5230 5240 5250 5260 5270 5280 5290

5300 5310 5320 5330 5340 5350 5360 5370

5380 5390 5400 5410 5420 5430 5440 5450

5460 5470 5480 5490 5500 5510 5520 5530

5540 5550 5560 5570 5580 5590 5600 5610

5620 5630 5640 5650 5660 5670 5680 5690

5700 5710 5720 5730 5740 5750 5760 5770

5780 5790 5800 5810 5820 5830 5840 5850

5860 5870 5880 5890 5900 5910 5920 5930

5940 5950 5960 5970 5980 5990 6000 6010

6020 6030 6040 6050 6060 6070 6080 6090

6100 6110 6120 6130 6140 6150 6160 6170

6180 6190 6200 6210 6220 6230 6240 6250

6260 6270 6280 6290 6300 6310 6320 6330

6340 6350 6360 6370 6380 6390 6400 6410

6420 6430 6440 6450 6460 6470 6480 6490

6500 6510 6520 6530 6540 6550 6560 6570

6580 6590 6600 6610 6620 6630 6640 6650

6660 6670 6680 6690 6700 6710 6720 6730

6740 6750 6760 6770 6780 6790 6800 6810

6820 6830 6840 6850 6860 6870 6880 6890

6900 6910 6920 6930 6940 6950 6960 6970

6980 6990 7000 7010 7020 7030 7040 7050

7060 7070 7080 7090 7100 7110 7120 7130

7140 7150 7160 7170 7180 7190 7200 7210

7220 7230 7240 7250 7260 7270 7280 7290

7300 7310 7320 7330 7340 7350 7360 7370

7380 7390 7400 7410 7420 7430 7440 7450

7460 7470 7480 7490 7500 7510 7520 7530

7540 7550 7560 7570 7580 7590 7600 7610

7620 7630 7640 7650 7660 7670

2500 2550 2600 2610 2620 2630 2640 2650

1780 1790 1800 1810 1820 1830 1840
ACCTAAAGGATCAACAGCTCTGTGGGATTGTTGGGGTTGCTCTGGAAACTCATTGTCACCACTGCTGTGCCTT
|| ||||| || || ||| | ||||| || | | || | ||| |||| ||| ||
ACTTAAGGAGCAGGAGCTGAATACCTTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTACCAT
2660 2670 2680 2690 2700 2710 2720

1850 1860 1870 1880 1890 1900 1910
GG----AATGCTAGTTTGSAGTAATAAATCTCTGGAACAGATTGGAATAACATGACCTGGATGGAGTGGGACA
|| ||||| ||| ||| | ||||| || || || | ||| ||||| |||||
GGCCAAATGCAAT-----CTAGCACCAGACTGGAACA-AT---GA-TA-----CTTGGCAAGAGTGGGAGC
2730 2740 2750 2760 2770

1920 1930 1940 1950 1960 1970 1980
GAGAAATTAAAC-AATTACACAAGCTTAATACATTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAGAAT
|| | ||| || || | || ||| || | ||||| | |||| | ||||| |||||
GAAAGGTTGACTTCTTGGAGGAAAATATAACA-GCCCTCCTAGAAGAGGCACAAATTCACAAGAGAGAAGAC
2780 2790 2800 2810 2820 2830 2840 2850

1990 2000 2010 2020 2030 2040 2050
GAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGG-AATTGGTTTAACATAACAAATTGGCTGTG
| ||||| | ||| ||| |||| | || | ||| ||||| ||| | | |||||
ATG)ATGAATTACAAAAGTTAAATAGCTGGG-ATGTGTTTGGCAATTGGTTTGACCTTGCTTCTTGGATAAA
2360 2370 2380 2390 2900 2910 2920

2060 2070 2080 2090 2100 2110 2120
GTATATAAAATATTCA--TAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGTCTGTACTTTCT
||||| |||| | ||| |||| ||| ||| | ||| ||||| ||| | |||
GTA)AT-ACAATATGGAATTTATG-TAGTTGTAGGAGTAATACTGTTAAGAATAGTGATCTATATAGTACAA
2930 2940 2950 2960 2970 2980 2990

2130 2140 2150 2160 2170 2180
ATAGTGAATAGAGTT-AGGCAGGATATTCACCA---TTATCGTTTCAGACCCAC---CTCCCA----AC--
|| | || ||||| |||| ||| ||| ||| ||| ||| ||| ||| ||| |||
ATGCTAGCTA-ATTAAAGGCAGGGGTATAGCCAGTGTTC-TCCCCAACCTCTTATTTCCAGTAGACTC
3000 3010 3020 3030 3040 3050 3060

2190 2200 2210 2220 2230 2240
---CCCAAGGAGACCGGACA-GGCC--C---GAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACA
|| | ||||| || ||| | ||||| ||||| ||| ||||| ||| ||| |||
ATACCCAACAGGACCGGCACTGCAACCCAGAGAGAGGCAAGAAGGAGACGGTGGAGAAGGCGGTGGCAACA
3070 3080 3090 3100 3110 3120 3130

2250 2260 2270 2280 2290 2300
GATC----GATTGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGC--CTTG---TGC
|| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
GCTCCTGGCCTTGGCAGATAGAAATATATTCATTTCTGATCCG---CCAAGTATACGCTCTTGACTTGG
3140 3150 3160 3170 3180 3190 3200

2310 2320 2330 2340 2350 2360 2370
CTCTTCAGCTACACCGCTTGAGAGACTTACTCTTGATTGTA-ACGAGGATTGTGGAAGTTCTGGGACGCAG
|| ||||| | || ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CTATTCAGCAA-----CTGCAGAACCTTGCTATCGAGAGCATACCA-GATCCTCCAACCAATACTCCAGAG
3210 3220 3230 3240 3250 3260

2380 2390 2400 2410 2420 2430 2440
GGGTTGGGAGGCGCT-CAAA-TATTGGTGGAAATCTCTACAGTATTGGAGTCA-GGAAGTAAAG
| | | |||| | || ||| | | ||||| ||| ||| ||| ||| ||| |||
GCTCTCTGCGACCTACGAGGATTCGAG--AAGTCCT-CAGGACTGAAGTACCTACCTACAA
3270 3280 3290 3300 3310 3320 X

6. KUNZ-158-CL33. SEQ

RESIMVXX Simian immunodeficiency virus STLV-III(MAC) provir

ID RF01VMXX standard: DNA; 9646 BP.
 XX
 AC Y00277:
 XX
 DT 17-SEP-1987 (annotation)
 XX
 DE Simian immunodeficiency virus STLV-III(MAC) proviral genome
 XX
 KW art gene; env gene; envelope protein; gag gene; genome;
 KW overlapping genes; pol gene; tat gene.
 XX
 OS Simian immunodeficiency virus
 OC Viridao; ss-RNA enveloped viruses; Retroviridae.
 XX
 RN [1] (bases 1-9646)
 RA Chakrabarti L. ;
 RT ;
 RL Submitted (18-AUG-1987) to the EMBL Data Library.
 XX
 RN [2] (bases 1-9646)
 RA Chakrabarti L. , Guyader M. , Alizon M. , Daniel M. D. ,
 RA Desrosiers R. C. , Tiollais P. , Sonigo P. ;
 RT "Sequence of simian immunodeficiency virus from macaque and its
 RT relationship to other human and simian retroviruses";
 RL Nature 323:543-547(1987).
 XX
 CC #source: library=infected HUT78 cells; clones=lamba SIV1;
 XX
 FH Key From To Description
 FH
 FT SITE 1 175 LTR (R)
 FT RPT 1 318 5' LTR
 FT SITE 176 318 LTR (U5)
 FT SITE 488 2068 ORF (gag)
 FT CDS 551 2068 gag gene product (AA 1-506)
 FT CDS 1726 4933 pol gene product (AA 1-1056)
 FT SITE 4793 5467 ORF (g)
 FT CDS 4826 5467 g gene product (AA 1-214)
 FT SITE 5295 5633 ORF (X)
 FT CDS 5298 5633 X gene product (AA 1-112)
 FT CDS 5637 5639 R gene product (AA 1-101)
 FT SITE 5758 6083 ORF, exon 1 (tat)
 FT CDS 5788 6083 tat gene product (AA 1-99)
 FT (6083 is 2nd base in codon)
 FT SITE 6011 6083 ORF, exon 1 (art)
 FT CDS 6014 6083 art gene product (AA 1-23)
 FT (6083 is 1st base in codon)
 FT SITE 6067 8735 ORF (env)
 FT IVS 6084 8300 intron I (tat and art)
 FT (8301 is 3rd base in codon)
 FT CDS 6080 8735 envelope protein (AA 1-882)
 FT SITE 8247 9551 ORF, exon 2 (art)
 FT SITE 8251 9394 ORF, exon 2 (tat)
 FT SITE 8298 8300 in-frame stop codon (env)
 FT CDS 8301 9394 tat gene product (AA 100-130)
 FT CDS 2301 9551 art gene product (AA 24-108)
 FT (8301 is 2nd base in codon)
 FT SITE 9560 9357 ORF (F)
 FT CDS 9572 9357 F-gene product (AA 1-262)
 FT SITE 9857 9471 LTR (U3)
 FT RPT 9857 9646 3' LTR
 FT PRM 9363 9377 enhancer-like sequence
 FT PRM 9382 9391 enhancer-like sequence
 FT SITE 9392 9402 conserved sequence
 FT SITE 9404 9412 conserved sequence
 FT SITE 9414 9424 conserved sequence

Sequence 3646 BP: 3250 A; 1817 C; 2387 G; 2152 T; 0 other;

Y	6130	6130	6140	6150	6160	6170	6180
---	------	------	------	------	------	------	------

6250

6320

6390

6460

6

6590

6730

[illegible]

TACTTCTGTTATCCAAAGAGTGTTGTGACAAAAGAT-TATTGGGATGCTATT-AGATGTAGTATTGTGCACCT
6740 6750 6760 6770 6780 6790 6800

680 690 700 710 720 730 740
GCTCGTTTTTCGATTCTAAAAATGTAATAATA-AGACGTTCAATGGAACAGGACCA--TGTACAAATGT-CAG
| | | | | | | | | | | | | | | | | | | | | |
CCAGGTTATGCTTTGCTTAGATGTAATGACACAAATTATTTCAGGCTTTATGCCTAACTGTTCTAAGGTAGTG
6810 6820 6830 6840 6850 6860 6870

750 760 770 780 790 800 810
CACAGTACAATGTACACATGGA--AT-TAGGC-CAGTAGTATCAACTCAACTGCTGTTGAATGGCAGTCTAG
| | | | | | | | | | | | | | | | | | | | | |
GTCCTTTC-ATG--CACAGGGATGATGGAGACACAG-ACT-TCTACTTGGTTTCGGTTTAATGGAAGTAGAG
6880 6890 6900 6910 6920 6930 6940

820 830 840 850 860 870 880
CAGAAGA-AGAGGTAGTAATTAGATCTGCCAATTTACAGACAATGCTAAAACCATAA-TAGTACAGCTGAA
| | | | | | | | | | | | | | | | | | | | | |
CAGAAAATAGAACCTATATTIA---CTGGC-ATGGTAGAGATAA---TAGGACTATAATTAGTCTAAAT-AA
6950 6960 6970 6980 6990 7000

890 900 910 920 930 940 950
CCAAT---CTGTAGAAATTAATTGTACAAGACCCAAACAATACAAGAAAAAGTATCCGTATCCAGAGGGG
| | | | | | | | | | | | | | | | | | | | | |
GCATTTATAATCTAACAAATGAATGTAGAAGA-CCAGGA--AAT--AAG--ACAGTTT---TA-CCA---GTC
7010 7020 7030 7040 7050 7060

960 970 980 990 1000 1010
ACCA--GGGAGAGCATT--TGTTCACATAGGAAAAATAGGAAATATGAGAC--AAGCA--CATTGT-AACAT
| | | | | | | | | | | | | | | | | | | | | |
ACCATTTATGCTCTGCATTGGTTTTCCACT--CACAAACCAGTCAATGAGAGGCCAAAGCAGGCATGGTGTAGGT
7070 7080 7090 7100 7110 7120 7130

1020 1030 1040 1050 1060 1070
TAGTAGAGCAAATGCAATGCCACTTTAAACAGAT--AGCTAGCAAATT----AAGAGACA---AT-TTG
| | | | | | | | | | | | | | | | | | | | | |
T--TGGAGGAAATTTGGAGGAGGCAATAAAGAGGTGAAGC-AGACCATTGTCAAACATCCCAGGTATACTG
7140 7150 7160 7170 7180 7190 7200

1080 1090 1100 1110 1120 1130 1140
GAAATAATAAAAG-AAT--AATC--TTTAAAGCAATCCT-CAGGAGGGGGACCCAGAAATTGTAACGCACA-GT
| | | | | | | | | | | | | | | | | | | | | |
GAACTAACNATACTGATAAAATCAATTTGACGGCTCCTAGAGGAGGAGATCCGGAA--GTTACCTTCATGT
7210 7220 7230 7240 7250 7260 7270

1150 1160 1170 1180 1190 1200
---TTTAATTGTGGAGGGGAATTTTTCTACTGTAATTCAACACAACCTGTTTAATAGTACTTGGTTTA--ATAG
| | | | | | | | | | | | | | | | | | | | | |
GGACAAATTGCAAGAGGAGAGTTTCTCTACTGTAA---AA-TGAATTGGTTTCTA-AATTGGGTAGAAGATAG
7280 7290 7300 7310 7320 7330

1210 1220 1230 1240 1250 1260 1270 1280
TACTTGGAGTACTGAAGGGTCAAAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCAGAATAAAACA
| | | | | | | | | | | | | | | | | | | | | |
GAGTCTAACTAOCACAGAGCCAAAGGA-ACGGCATAAAAGG-----AATTAC-GTACCATGT CATATTAGACA
7340 7350 7360 7370 7380 7390 7400

1290 1300 1310 1320 1330 1340 1350
ATTTATAACCATGTGGCAGGAAGTAGGAAAAGCAATG--TATGCCCTCCCATCAGCGGACAAATTAGATGT
| | | | | | | | | | | | | | | | | | | | | |
AATCATCAGCACTTGGCATAAGGTAGGCAAA--AATGTTTATTTGCCCTCCAGAGAGGGAGACCTCACGTGT
7410 7420 7430 7440 7450 7460 7470

1360 1370 1380 1390 1400 1410 1420
TCATCAAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAACAACAATGGGTCCGAG-ATCTTCAG
| | | | | | | | | | | | | | | | | | | | | |

2130 2140 2150 2160 2170
TCTATAGTGAATAGAGTTAGGCAGGGATATTCACCA---TTATCGTTTCAGACCCACCT-----C

```

CAATAGCTAGCTAGCTTAAAGACAGGGGTATAGCCAGTGTCTC-TTCCCCACCCCTCTTATTTCCAGTAGAC
      8240      8250      8260      8270      8280      8290      8300

2180      2190      2200      2210      2220      2230      2240
CCAACCCCGAGGAGACCCGAC-AGGCC--C-----GAAGGAATAGAGAGAGAGGTGGAGAGAGAGACAGAGA
||| ||| | ||| ||| | ||| | |||| | | ||| ||| ||||| ||| | |
CCATACCCACAGGATCGGGCTCTGCCAACCAAGAAGGCCAAAAAGGAGACGGTGGAGGCAGCGGTGGCAA
      8310      8320      8330      8340      8350      8360      8370

      2250      2260      2270      2280      2290      2300
CAGATC-----CATTGGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGC--CTTG---T
||| || | ||| | ||| |||| | ||| ||| | ||| ||| ||| ||| ||| |||
CAGCTCCTGGCCTTGGCAGATAGAAATATATTCATTTTCTGATCCG---CCAACTGATACGCCTCTTGACTT
      8380      8390      8400      8410      8420      8430      8440

2310      2320      2330      2340      2350      2360      2370
GGCTCTTCAGCTACCAACCGCTTGAGAGACTTACTCTTGATTGTA-ACGAGGATTGTGGAACCTCTGGGACGC
| | ||||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |
GGCTATTCAACAA-----CTGCAGAACCTTGCTATCGAGAGCATACCA-GATCCTCCAACCAATATTCCAG
      8450      8460      8470      8480      8490      8500

2380      2390      2400      2410      2420      2430      X
AGGGGGTGGGAAGCCCT-CAATATTGGTGGAAATCTCCTACAGTATTGGAGTCA-GGAACATAAG
||| | | |||| | ||| ||| |||| | ||| ||| ||| ||| ||| |||
AGGCTCTCTGCACCCCTACCGAGAAATTCGAGAA-GTCCT-CAGGCTTGAACCTACCTACCA
      8510      8520      8530      8540      8550      8560      8570

```

7. KUNZ-158-CL33. SEQ

RESIVAXX Simian immunodeficiency virus STLV-III(AGM) provir

```

ID  RESIVAXX  standard; DNA; 9264 BP.
XX
AC  Y00295;
XX
DT  28-SEP-1987 (annotation)
XX
DE  Simian immunodeficiency virus STLV-III(AGM) proviral genome
XX
KW  env gene; envelope glycoprotein; gag gene; genome;
KW  long terminal repeat; overlapping genes; pol gene; protease;
KW  B gene; reverse transcriptase; sor gene; transfer RNA;
KW  transfer RNA-Lys; unidentified reading frame.
XX
OS  Simian immunodeficiency virus
OC  Viridae; ss-RNA enveloped viruses; Retroviridae.
XX
RN  [1] (bases 1-9264)
RA  Reitz M.;
RT  ;
RL  Submitted (17-JUL-1987) on tape to the EMBL Data Library by:
RL  Reitz M., LTCB, NIH, Bldg 37 Room 6C09, Bethesda md 20892, USA.
XX
RN  [2] (bases 1-9264)
RA  Franchini G., Gallo R.C., Guo H.G., Gurgo C., Callatti E.,
RA  Fargnoli K., Hall L., Wong-staal F., Reitz M.S.;
RT  "Sequence of simian immunodeficiency virus and its relationship to
RT  the human immunodeficiency viruses";
RL  Nature 328:539-543(1987).
XX
CC  *source: library=EMBL-3; cell line=infected K6W
XX
FH  Key          From      To          Description
FH
FT  SITE         1        178        R-region of 5'-LTR
FT  SRT          1        303        5'-long terminal repeat

```

XX

Gaps

200

```

      280      290      300      310      320      330
--GAAAATTTTAAACATGTGGAAAA--ATGA-CAT--GGTAGAAC---AGATGCATGAGGATATAATCAG-TT
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
CCTCAATAAAGCCTTGTGTGTAAGATTATCCCCATTATGCATTACTATGAGATGCAATAAAAGTGAGACAGATA
   6020      6030      6040      6050      6060      6070      6080

      340      350      360      370      380      390
TAT-EGGAT---CA-AAGCCTAAA-GCCATGTGTGTAAGATTAAACCC-AC-TCTGTGTTAGT-TTAAAGTGCA
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
GATCGGGATTGACAAATCATCAACAACAATAACAACAGCAGCACCACCAACATCAGCACCAGTATCAGAAAAA
   6090      6100      6110      6120      6130      6140      6150

```

400 410 420 430 440 450 460
CTGATTTTGGGGAATGCTACTAATAACCAATACTAG-TAATACCAATAGTAGTAGCGGGGAA-ATGA-TGATGG
| | | | | | | | | | | | | | | | | |
TASACATGGTCAATGAGACTAGTTCTTGTA-TAGCTCAGAATAATTGCACAGGCTTGGAACAAGAGCAAATG
6160 6170 6180 6190 6200 6210 6220

470 480 490 500 510 520 530
AGAAAGGAGAGATAAANAAGTGTCTTTCAATATCAGCACAGNATAAGAGGTAAAGGTGCAGAAAG--AATA
| | | | | | | | | | | | | | | | | |
ATAAGCTGTAAATTCACCATGACAGGGTTAAAAAGAG-ACAAGACAAGGAGTACAATGAAACTTGGTACTC
6230 6240 6250 6260 6270 6280 6290 6300

540 550 560 570 580 590 600
TGC---ATTT-TTTTTATAAAGTTGATATAATACCAATAGATAATGATACTACCAGCTATACGTTGACAAGTTG
| | | | | | | | | | | | | | | | | |
TACAGATTGGTTTTGTGAAC--AAGGGAATAGCACT-GATAATGAAAGCAGATGCTACA---TGAATCACTG
6310 6320 6330 6340 6350 6360

610 620 630 640 650 660 670
TAACACCTCAGTCATTACACAGGCCCTGT-CCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCC
| | | | | | | | | | | | | | | | | |
TAACACTTCTGTTATCCAAGAGTC TTGTGACAAACAT-TATTGGGATACTATTAGATTTAGGTATTGTGCAC
6370 6380 6390 6400 6410 6420 6430

680 690 700 710 720 730 740
CGGCTGGTTTTGCGATTCTAAAATGTAATAATA-AGACGTTCAATGGAACAGGAC--CATGTACAAATGT-C
| | | | | | | | | | | | | | | | | |
CTCCAGGTTATGCTTTGCTTAGATGTAATGACACAAATTATTCAGGCTTTATGCCTAAATGTTCTAAGGTGG
6440 6450 6460 6470 6480 6490 6500

750 760 770 780 790 800
AGC/CAGTACAA/TGTACACATGGA--AT-TAGGC-CAGTAGTATCAACTCAACTGCTGTTGAATGGCAGTCT
| | | | | | | | | | | | | | | | | |
TGGTCTCTTC-A/G-- CACAGGGATGATGGAGACACAG-ACT-TCTACTTGGTTTGGCTTTAATGGAAGTAG
6510 6520 6530 6540 6550 6560 6570

810 820 830 840 850 860 870
AGCAGAGA-AGAGGTAGTATTTAGATCTGCCAATTTACAGACAATGCTAAAACCATAA-TAG--TACAGC
| | | | | | | | | | | | | | | | | |
AGCAGAAAATACAACTTATATTTA---CTGGC-ATGGTAGGGATAA---TAGGACTATAATTAGTTTAAATA
6580 6590 6600 6610 6620 6630 6640

880 890 900 910 920 930 940
TGAACCAATCTGAGAAATTAATTGTACAAGACCCAACACAACAATACAAGAAAAAGTATCCGTATCCAGAGGG
| | | | | | | | | | | | | | | | | |
AGTATTATAATCTAACAATGAAATGTAGAGA-CCAGGA--AAT--AAG--ACAGTTT---TA-CCA---GT
6650 6660 6670 6680 6690

950 960 970 980 990 1000 1010
GACCA--GGGAGAGCAATT--TGTTACAAT-AGGAAAAATAGGAAATATGAGACAAGCA--CATTGTAACATT
| | | | | | | | | | | | | | | | | |
GACCATTATGTCTGATTTGGTTTTTCCACTCACACCCTACTGATAGGCCA-AAGCAGGCATGGT---GTT
6700 6710 6720 6730 6740 6750 6760

1020 1030 1040 1050 1060 1070
AG--TAGAGCAAAATGCAATGCCACTTTAAACAGAT-AGCTAGCAAATT----AAGAGAACA---AT-TTG
| | | | | | | | | | | | | | | | | |
GGTTTGGAGGAAATGGAAGATGCAATAAAAGAGGTGAACAGACCATTGTCAAACATCCCAGGTATACTG
6770 6780 6790 6800 6810 6820 6830

1080 1090 1100 1110 1120 1130 1140
GAATAATAGAAC-AAT--AATC--TTTAAGCAATCCT-CAGGAGGGGACCCAGAAATTGTAACGCACA-GT
| | | | | | | | | | | | | | | | | |
GAADTAACATATCTGATAAATCAATTTAGCTCTCCTGGAGGAGGAGATCCGGAA---GTTACCTTCATGT
6840 6850 6860 6870 6880 6890 6900

OS RDD isolate
 OC Viridae: ss-RNA enveloped viruses; Retroviridae.
 XX
 RN [11 (bases 1-9671)
 RA Allison M.
 RT ;
 RL Submitted (03-JUN-1987) on tape to the EMBL Data Library by:
 RL Marc Allison, Unite d'Oncologie virale, and CNRS UA1157, Institut
 RL Pasteur, 25 rue du Dr Roux, 75724 Paris CEDEX 15, France.
 XX
 RN [21 (bases 1-9671)
 RA Guyader M., Emerman M., Sonigo P., Clavel F., Montagnier L.,
 RA Allison M.
 RT "Genome organization and transeactivation of the human
 RT immunodeficiency virus type 2";
 RL Nature 326:652-668(1987).
 XX
 RN [31
 RA Clavel F., Guyader M., Guetard D., Salle M., Montagnier L.,
 RA Allison M.
 RT "Molecular cloning and polymorphism of the human immune deficiency
 RT virus type 2";
 RL Nature 324:681-685(1987).
 XX
 FH Key From To Description
 FH
 FT SITE 1 173 R region
 FT RPT 1 253 LTR
 FT SITE 1 9671 HIV-2 RNA
 FT corresponding to integrated
 FT proviral DNA
 FT SITE 174 298 U5 region
 FT SITE 303 320 primer binding site
 FT CDS 546 2111 gag protein
 FT CDS 1829 4935 pol protein
 FT SITE 4513 4626 polypurine tract 2
 FT CDS 4563 3512 q protein
 FT CDS 5582 5596 r protein
 FT CDS 5645 6140 tat protein part 1
 FT (6140 is 2nd base in codon)
 FT CDS 6071 6140 art protein part 1
 FT (6140 is 1st base in codon)
 FT CDS 6147 8720 env protein
 FT CDS 8307 8400 tat protein part 2
 FT (8307 is 3rd base in codon)
 FT CDS 8307 8536 art protein part 2
 FT (8307 is 2nd base in codon)
 FT CDS 8557 9324 f protein
 FT SITE 8925 8935 polypurine tract 1
 FT SITE 8942 9487 U3 region
 FT RPT 8942 9671 LTR
 FT PRM 9329 9339 core enhancer sequence
 FT PRM 9401 9416 core enhancer sequence
 FT SITE 9420 9427 pot. SP1 factor binding site
 FT SITE 9428 9437 pot. SP1 factor binding site
 FT SITE 9438 9443 pot. SP1 factor binding site
 FT PRM 9465 9470 TATA-box
 FT SITE 9493 9671 R region
 FT SITE 9643 9654 pot. polyA signal
 FT POLYA 9671 9671 polyA site
 XX
 SQ Sequence 9671 BP; 3314 A; 1973 C; 2401 G; 1983 T; 0 other;

Initial Score = 306 Optimized Score = 1185 Significance = 0.00
 Residue Identity = 53% Matches = 1395 Mismatches = 888
 Gaps = 323 Conservative Substitutions = 0

X 10 20 30 40 50 60
ATGAG---AGTGAAGAGAAATATCAGC-ACTTGTGGAGATGGGGGTGGAAATGGGGGCACCATGCTCCTTGG
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACCAAGACAAATTAATATGATGATCAGCTGCTTATTGCCATTTTATT--AGCTAGTGC---TTG---CTTAG
X 6140 6150 6160 6170 6180 6190

70 80 90 100 110 120 130
GATATTG-----ATG-ATCTGTAGTGCTACAGAAAAATTGTG-GGTCACAGTCTATTATGGGGTACCTG
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
TATATTGCACCCATATGTACTGT-TTCTATGCGGTACCCACGTGGAAAAATGCAACCATTCCTCTTT
6200 6210 6220 6230 6240 6250 6260

140 150 160 170 180 190
TGTEGAAGGAAGCAACCACTCTATTTTGTG---CAT-CAG---ATGCTAAAGCAT-ATGATACAGAGGT
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
TGATCAA-CCAGAAATAGGGA---TACTTGGGGAACCATACAGTGCTTGCTGACAATGATGATTATCAGG-
6270 6280 6290 6300 6310 6320 6330

200 210 220 230 240 250 260
ACAT AATGTTTG---GGCCACACATGCTGTGTACCCACAGACCCCAAC-CCACAAGAAGTAGTATTGGTAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AAATAG-CTTTGATGTAAACAGAGCTTTTG-ATGCATGGAATAATACAGTAACAGAACAGCAATAGAAGA
6340 6350 6360 6370 6380 6390 6400

270 280 290 300 310 320 330
TG--TGACAGAAAATTTTAACATGTGGAAAAATGACATG-GTAGAACAGATGCATGAGGATATAATCAGTTT
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
TGTCTGGCATCTATTGAGACAT---CAATAAAACCATGTGTCAAACCTAACACCTTTATGTGT-AGCAATGA
6410 6420 6430 6440 6450 6460 6470

340 350 360 370 380 390 400
ATGGGATCAAAACCTAAAGC---CATGTGTAAAATTAACCCCACTCTGTGTTAGTTTAAAGTGCCTGATTG
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AATGCAGCAGCAGAGAGAGCAGCAGCAGGAA---CAACACAAC-CTCAAAGAGCACAA---GCACAACCACA
6480 6490 6500 6510 6520 6530

410 420 430 440 450 460
GGGAATGCTACTAATACCAATA-CTAG---TAA-TACCAATAGTAGTAGCG---GGGAATGA--TGATGGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACCACACCCAC---AGACCAGGAGCAAGAGATAAGTGAGGATACTCCATGCGCAGCGCAGACAACCTGCT-CA
6540 6550 6560 6570 6580 6590 6600

470 480 490 500 510 520 530
GAAGAGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAGNATAAGAGG-TAAGGTGCAGA-AAGAATAT
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
GGATGGGAGAGGAAGAAACGATC---AATTECCAGTTCAA-TATGACAGGATTAGAAAGAGATAAGAAAAA
6610 6620 6630 6640 6650 6660 6670

540 550 560 570 580 590 600
GCATTTTATTATAACTTGATATAATACCAATAGATAATGATACTACCAGCTATACGTTGACAAGTTGTAAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACAGTATAAT-GAATCATGGTA----CTCAAAGATGTGGTTTGTGAGACAAATAATAGCACAA---ATCAG
6680 6690 6700 6710 6720 6730

610 620 630 640 650 660 670
ACCCAGTTCATTACACAGGCTGTCCAAAGGTATCCTTTGAGCCCAATCCCATACATTATTGTGCCCCGGCT
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACC-CAGT-GTTACATGAACCATTTGC-AACACATC-----AGTCA--TCACAGA-ATCA-TGTGACAAG--C
6740 6750 6760 6770 6780 6790

680 690 700 710 720 730 740
GGTTTTCGATTTCTAAATGTAATAA-AGACGTTCAATGGAACAGGACC-ATGTACA---AATGTCAGCA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACTATTGGGATGCTATAAGGTTTAGATACTGTGCACCAACCGGGTTATGCCCTATTAAAGATGTAATGATACCA
6800 6810 6820 6830 6840 6850 6860

750 760 770 780 790 800
GASTA--CA----ATGTACACA--TGGAAATTAGGCCAGTAGTATCAAC--TCAAC--TGCTG-TTGAATGGCA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
-ATTATTTCAGGCTTTGCACCCAAC TGTCTTAAG--TAGTAGCTTCTACATGCACCAGGATGATGGAAACGCA
6970 6980 6990 6900 6910 6920 6930

810 820 830 840 850 860 870
GTCTAGCAGAGAGAAGAGGTAGTAATTAGATCTGCCAATTTCCACAGACAATGCTAAAACCA--TAATAGTACA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AACTTCCACATG---GTTTTGGCTTTA-A--TGGCACTAGAGCAGAGAATAGAACATATATCTATTGGCATG
6940 6950 6960 6970 6980 6990 7000

880 890 900 910 920 930 940
GCTGAACCAATCTGTAGAAATT TAATTGTACAAGACCCCAACAACAATACAAGAA--AAAGTATCGTATCCAG
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCAGAGATAA---TAGAACT--ATCAT-CA-GCTTAAACA--AATATTATAATCTCAGTTTGCATTGTAAG
7010 7020 7030 7040 7050 7060

950 960 970 980 990 1000 1010
AGGGGACCCAGGGA--GAGCATT TTGTAC-AATAGGAAAAAT---AGGAAATATGAGACAAGCACATTGTAAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGG---CCAGGGGATAAGACAGTGAACAACAAT AATGCTTATGTCAGGACATGTGTTTCACTCCCCT---AC
7070 7080 7090 7100 7110 7120 7130

1020 1030 1040 1050 1060 1070
ATTAGTAGAGCAAAATGCAATGCGACTTTTAAACA---GATAG--CTA--GCAAATTAAGAGA-ACAAT-T
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
--CAGCCGATCAATA---AAGAGCCAGACAAGCATGGTGTGCTGGTTCAAAGGCAATGGAAAGACGCCATGC
7140 7150 7160 7170 7180 7190

1080 1090 1100 1110 1120 1130 1140
TGGAAAT-AATAAAACAATAATCTTTAAGCAATCCTCAGG---AGGGGACCCAGAAATTGT AACGCACAGTT
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGGAGGTGAAGGAAGGCT---TGCAAAACATCC-CAGGTATAGAGGAACC---AAT-GACACAAGGAATA
7200 7210 7220 7230 7240 7250 7260

1150 1160 1170 1180 1190
TTAATTGTGGAG-----GGGAATTTTC-TAC-----TGTA--ATTCA---ACAC-AACTGTTTAATAGTAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
TTAGCTTTTCAGCGCGCAGGAAAGGCTCAGACCCAGAAAGTAGCATACATGTGGACTAACTGCAGAGGAG-AG
7270 7280 7290 7300 7310 7320 7330

1200 1210 1220 1230 1240 1250 1260
TTGGTTTAAATAGTACTTGGAGTACTGAAGGGTCAAAT-AACACTGAAGGAAGTGACA---CAATCACACTCC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
TTTCTCTACTGCAACAT-GACT--TGGTT CCTCAATTGGATAGAGAA-TAAGACACACCGCAATTATGC-AC
7340 7350 7360 7370 7380 7390

1270 1280 1290 1300 1310 1320 1330
CATGCAGAAATAAACCAATTTATAAACATGTGGCAGGAAGTAGGAAAAGCAATG--TATGCCCTCCCATCAG
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
CGTGCCATATAAAGCAATAATTAACACATGGCATAAGGTAGG--GAGAAATGTATATTTGCCTCCCA----
7400 7410 7420 7430 7440 7450 7460

1340 1350 1360 1370 1380 1390 1400
CGSACAAA TATATGTTTCATCAAATATTACAG--GGCTGC-TATTAACAAGAGATGG--TGGTAATAACAAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGGA-AGGGAAGCTTCTCTGCAACT-CAACAGTAACCAGCATAATTGCTA-ACATTGACTGGCAA-AACAAT
7470 7480 7490 7500 7510 7520 7530

1410 1420 1430 1440 1450 1460
AATGGGTCCGASAT---CTTCAGACCTGGAGGAGG-AGA--TATGAGGGACA-ATTGGAGAAGTGAATTATA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AATCAGACAGAGATTACCTTTAGTGC-AGAGGTGGCAGAACTAT-----ACAGATTGGAGTTGGGAGAT-TA
7540 7550 7560 7570 7580 7590

[illegible]

```

2230      2240      2250      2260      2270      2280      2290
GGAGAGAGAGACAGGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCTGCGG
||| | | | ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GGAGGCAACGGTGGAGACAGATACTGGCCCTGGCCGATAGCATATATA-CA-TTTCCTGATCCG-CCAGCTG
      8370      8380      8390      8400      8410      8420      8430

2300      2310      2320      2330      2340      2350      2360
AGCCCTGTTCCTCTT---CAG-CTA-CCA-CCGCTTGAGAGACTTACTCTTGATTGTAAC--GAGGATTGTG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
A---TT-CGCTCTTGACCAAGACTATACAGCATCTGCAGGGACTTACTATCCA--GGAGCTTCCTGACCCCTC
      8440      8450      8460      8470      8480      8490

      2370      2380      2390      2400      2410      2420      2430
GAAGT--TCTGGAGCGCAGGGGGTGGGAAGCCCTCAAATA-TTGGTGGGATC-TCCTACAGTATTGGAGTCA
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CAATCATCT---AC-CAGAAATCTCAGAGACTGGCTGAGACTTAGAACAGCCCTTCTTGCAATATGGGTGCGA
8500      8510      8520      8530      8540      8550      8560

      X
--GGAAGTGA--AAG
||| | | | | |
GTGATCCAGGAG
8570      8580

```

9. KUNZ-158-GL33 SEB

RESIV251 Simian Immunodeficiency virus (Mac251) envelope ge

```

ID RESIV251 standard? DNA; 1142 BP.
XX
AC X06879; Y00264;
XX
DT 23-JUN-1988 (acc# added)
DT 26-MAY-1989 (annotation)
XX
DE Simian Immunodeficiency virus (Mac251) envelope gene DNA (part.)
DE integrated copy
XX
KW env gene; envelope gene.
XX
OS Simian immunodeficiency virus
OC Viridae; ss-RNA enveloped viruses; Retroviridae.
XX
RN [1] (bases 1-1142)
RA Kestler H.W.;
RT ;
RL Submitted (25-FEB-1988) to the EMBL Data Library by:
RL Kestler H.W.; Harvard Medical School, New England Regional Primate
RL Research Center, Department of Microbiology, One Pine Hill Drive,
RL Southborough, Mass. 01772; USA.
XX
RN [1]
RA Kestler H.W.; Li Y.; Naidu Y.M.; Butler C.V.; Ochs M.F.;
RA Jansen D.; King N.W.; Daniel M.D.; Desrosiers R.C.;
RT "Comparison of Simian immunodeficiency virus isolates";
RL Nature 331:618-622(1988).
XX
CC source: strain=Macaca mulatta 251(host); clone=lambd SIV 251;
XX
FH Key From To Description
FH
FT SITE 1 1142 put. env gene
FT (1 is 2nd base in codon)
FT SITE 1026 1029 in frame stop codon
XX

```

SQ Sequence 1142 BP: 322 A: 208 C: 273 G: 231 T: 108 other:

Initial Score = 259 Optimized Score = 565 Significance = 0.00
Residue Identity = 51% Matches = 622 Mismatches = 488
Gaps = 88 Conservative Substitutions = 0

```

1140      1150      1160      1170      1180      1190      1200
ACAGTTTAAATTGTGAGGGGGAATTTTCTACTGTAAATTCACACAACTGTTTAAAGTACTTGGTTTA--A
|      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
TGTGACAAATTCACAGAGGAGAGTTCTCTACTGTAA---AA-TGAATTGGTTTCTA-AATTGGGTAGAGGA
X      10      20      30      40      50      60

1210      1220      1230      1240      1250      1260      1270
TAGTACTTGAAGTACTGAAGGGTCAAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCAGAATAAA
|||  |  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
TAGCGATGTAACTAGCCAGAGGCCAAAGGA-AC----GGCATAGAAGGAATTAC-GTGCCGTGTCATATTAG
70      80      90      100      110      120      130

1280      1290      1300      1310      1320      1330      1340
ACAATTTATGAACATGTGGCAAGAAAGTAGGAAAAGCAATG--TATGCCCTCCCATCAGCGGACAAATTAGA
|||  |  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
ACAATAATCAACACTTGCATAGAGTAGGCAA--AATGTTTATTTGCCTCCAAGAGAGGAGACCTCACG
140      150      160      170      180      190      200

1350      1360      1370      1380      1390      1400      1410
TGTTCATCAATATTACAGGCTCTCTATTACCAAGAGATGGTGGTAATAACAACATGGGTCCGAGATCTTC
|||  |  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
TGTAAGTCAAGAGTGAACAGTCTCATAGCAACATAGATTGGACTGATGGAAACCA---AACTAATATCACC
210      220      230      240      250      260      270

1420      1430      1440      1450      1460      1470      1480
AGAGCTG--GAGGAGT-AGATATGAGGACAAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAATT
|  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
ATGAGTGCAGGCTTGGCAGAACT--GTATCGATTGGAGTTGGGAGAT-TATAAT-----TAGTAGAGATN
280      290      300      310      320      330

1490      1500      1510      1520      1530      1540      1550
GAAGCATTATGAAGTACGACCCACCAAGGCAAGAG----AAGAGTGGT-GCA--GAGAGAAAAAGAGCAGT
|||  |  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
NNNNNNNNNNNNNNNNNNNNCCCCACAGATGTAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAG-AG-
340      350      360      370      380      390      400

1560      1570      1580      1590      1600      1610      1620
GAGGATAGGAGCTTCTGCTTGGGTCT-TGGAGCAGCAGGAAGCACTATGGGCGCACGGTCAATGACGC
|||  |  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
GAGTCTTTTCTAGAGTTCTTGGTTTCTCGCAACGGCAGGTTCTGCAATGGGCGCGGCGTCNNNNNNNN
410      420      430      440      450      460      470

1630      1640      1650      1660      1670      1680      1690
TGACGGTACGAGCCACACAAATTTATGTCTGGTATAGTGACAGCAGCAGAACAAATTTGCTGAGGGCTATTGAGG
|||  |  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
NNNCGGCTCAGTCCCGGACTTATTGGCTGGGATAGTGACGCAACAGCAACAGCTGTTGGACGTGGTCAAGA
480      490      500      510      520      530      540

1700      1710      1720      1730      1740      1750      1760
CGCAACAGCAGTCTGTTGCAACTCAGAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAA
|||||  |  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
GACACACAGAAATGTTTGGCACTGACCGTCTGGGGAACAAGAACCTCCAGACTAGGGTCACTGCCATCGAGA
550      560      570      580      590      600      610      620

1770      1780      1790      1800      1810      1820      1830      1840
GATACCTAAGGATCAACAGCTCTTGGGGATTGGGGTTGCTCTGGAAACTCATTTGCACCACTGCTGTGC
|||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
AGTACTTAAAGACCAAGCCGAGCTGAGTGTCTGGGGATGTGCGTTTAGACAAGTCTGCCACACTACTGTAC
630      640      650      660      670      680      690
```


[illegible]

FH

XX

SQ Sequence 308 BP: 85 A: 68 C: 87 G: 66 T: 0 other:

Initial Score = 194 Optimized Score = 298 Significance = 0.00
Residue Identity = 89% Matches = 299 Mismatches = 2
Gaps = 1 Conservative Substitutions = 0

2140 2150 2160 2170 2180 2190 2200 2210
TAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGG
|||||
TAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCCAATCCCGAGGGGACCCGACAGGCCCGAAGG
X 10 20 30 40 50 60 70

2220 2230 2240 2250 2260 2270 2280
AATAGAAGAGAGAGGTGGAGAGAGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTAT
|||||
AATAGAAGAGAGAGGTGGAGAGAGAGACAGAGACAGATCCATTGATTAGTGAACGGATCCTTAGCACTTAT
80 90 100 110 120 130 140

2290 2300 2310 2320 2330 2340 2350
CTGGGACGATCTGCGGAGCCTTGTGCTCTTCAGCTACCCACCGCTTGAGAGACTTACTCTTGATTGTAACGA
|||||
CTGGGACGATCTGCGGAGCC-TGTGCTCTTCAGCTACCCACCGCTTGAGAGACTTACTCTTGATTGTAACGA
150 160 170 180 190 200 210

2360 2370 2380 2390 2400 2410 2420
GGATTGTGGAACCTTCTGGGACGCAAGGGGGTGGGAAGCCCTCAAATATTGGTGGGAATCTCCTACAGTATTGGA
|||||
GGATTGTGGAACCTTCTGGGACGCAAGGGGGTGGGAAGCCCTCAAATATTGGTGGGAATCTCCTACAGTATTGGA
220 230 240 250 260 270 280

2430 X
GTCAGGAAGCTAAG
|||||
GTCAGGAAGCTAAG
290 300

